MICROBIOME OF TRADITIONAL CHINESE SOY SAUCE FERMENTATION BRINE

JOANITA BINTI SULAIMAN

DISSERTATION SUBMITTED IN FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTERS OF SCIENCE

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2014

ABSTRACT

The traditional Chinese soya sauce quality is determined by the corresponding metabolic behaviour of microbial population in the fermentation tank. In this study, the microbiome changes and functional capabilities throughout the traditional Chinese soya sauce brine fermentation process in a period of six months were investigated. Samples were obtained from soya sauce fermentation tank recurrently during the six months of the fermentation process. The metagenomes were analyzed using whole-genome shotgun (WGS) sequencing via next generation sequencing (NGS) on MiSeq (Illumina) platform. Alignment of the sequence reads with 16S rRNA and 18S rRNA genes revealed that soya sauce microbiome was predominated by members of the bacterial genus *Weissella* in the early stages and later succeeded by members of the fungal genus *Candida* at the end of the fermentation process. Upon analysis of the metabolic reconstructions displayed genetic potential with the majority of the genes were responsible for carbohydrate and energy metabolism besides the production of acid amino metabolism. Kyoto Encyclopedia of Genes and Genomes (KEGG) classification scheme using MetaGenome Analyzer (MEGAN) demonstrated a typical profile of heterotrophic fermentation of proteins and carbohydrates in which collaborated with the biochemical adjustments. This study provides new findings and insight of temporal changes in microbial successions over the six months of soya sauce brine fermentation and the biological processes within.

ABSTRAK

Kualiti kicap soya tradisional Cina ditentukan oleh komuniti mikrobial dan peranan metabolik yang wujud dalam tangki penapaian. Dalam kajian ini, perubahan populasi mikroorganisma dan keupayaan metabolik dalam proses penapaian selama enam bulan disiasat. Sampel diperolehi berkala dari tangki penapaian kicap soya dan dianalisa menggunakan *whole-genome shotgun* (WGS) *sequencing* secara *next generation sequencing* (NGS) dengan menggunakan mesin MiSeq (Illumina). Penjajaran dengan gen 16S rRNA dan 18S rRNA menunjukkan bahawa populasi kicap diterajui oleh genus bakteria *Weissella* pada peringkat awal dan kemudianya dikuasai oleh kulat dari genus *Candida*. Penilaian pembinaan metabolik mendedahkan potensi genetik bahawa majoriti daripada gen yang diperolehi adalah bertanggungjawab untuk metabolisme karbohidrat dan tenaga selain untuk metabolisme asid amino. Skim klasifikasi *Kyoto Encyclopedia of Genes and Genomes* (KEGG) menerusi *MetaGenome Analyzer* (MEGAN) menunjukkan profil ciri penapaian heterotrofik protein dan karbohidrat di mana berkait rapat dengan perubahan biokimia. Kajian ini menyediakan penemuan baru dan perubahan populasi mikroorganisma dalam tempoh enam bulan proses penapaian kicap tradisional Cina dan proses biologi di dalamnya.

ACKNOWLEDGEMENTS

I would like to extend my gratitude to Dr Chan Kok Gan for his continued supervision and support. I am indeed very grateful for his vision and foresight in which inspired and motivated me throughout this study.

I would like to express my appreciation to Gan H. M., Bernard-Lee K. B. and Ahmad Yamin for their valuable help and assistance in this study. I am also thankful to the late Chan Kun Lok for supplying the soya sauce samples and access to the soya sauce brine fermentation tanks.

I take this opportunity to record my sincere thanks to all my laboratory members who, directly or indirectly have assisted me in this study. I also thank my family for their unceasing support and encouragement.

Last but not least, I thank University of Malaya and the Ministry of Health (MOH) Malaysia for providing me an opportunity to complete my research study.

TABLE OF CONTENTS

PAGE

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

%	Percentage
<	Less than
µg	Microgram
µl	Microlitre
µm	Micrometre
µM	Micromolar
ATP	Adenosine triphosphate
BLASTN	Basic local alignment search tool using nucleotide query against nucleotide database
bp	Basepair
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
et al.	et alia (and others)
g	Gram
g	Gravity (relative centrifugal force)
Gb	Gigabase pair
h	Hour(s)
H_2O	Water
HCl	Hydrochloric acid
ISO	International Organization for Standardization
kb	Kilobase pair
L	Litre
M	Molar
mg/kg	Milligram per kilogram
min	Minute(s)
ml	Mililitre
mM	Milimolar
N	Normality
N/A	Not applicable

$Na_2EDTA.2H_2O$	Ethylenediaminetetraacetic acid disodium salt dehydrate
NaCl	Sodium chloride
$NaHPO_4$	Sodium phosphate
NaH_2PO_4	Sodium dihydrogen phosphate
NaOH	Sodium hydroxide
ng	Nanogram
ng/µl	Nanogram per microlitre
nM	Nanomolar
^oC	Degree Celcius
PCR	Polymerase chain reaction
pM	Picomolar
psi	Pound per square inch
rRNA	Ribosomal ribonucleic acid
s	Second(s)
SEM	Standard error of mean
SSU	Small subunit
U/µl	Unit per microlitre
V	Voltage
v/v	Volume per volume
v/w	Volume per weight
w/w	Weight per weight

CHAPTER 1.0

INTRODUCTION

The quality of soya sauce relies on the microbial population during its fermentation process (Hutkins, 2006; O'toole, 1997; Yong & Wood, 1974). The fermentation process of soya sauce involves two fermentation stages specifically the solid and liquid state fermentations. The typical method is by mixing the boiled soya beans with wheat flour, with additional inoculation of the fungi *Aspergillus*. As the mold develops and matures, it releases enzymes which catabolize the proteins present in the soya pastes. Upon maturation of the solid state fermentation, the paste known as *koji* will be placed in a fermentation tank for liquid state fermentation by addition of salt water (Hutkins, 2006; O'toole, 1997; Yong & Wood, 1974). Unlike other soya sauce production, the fermentation process of traditional Chinese soya sauce, both its solid and liquid states, do not involve defined inoculum. In traditional Chinese soya sauce fermentation processes, manufacturers depend solely on natural microbial selection to produce the end product. Currently, the usage of starter cultures under mass production fermenters fail to duplicate the whole microbial community of traditional fermentation. This leads to the production of less fragrant products (Zhao *et al*., 2009). Therefore, determining the whole microbial population involved in the traditional fermentation process may allow future improvement in production output whilst still maintaining the required soya sauce qualities.

Previous studies on soya based fermentation microbiome were done using PCR-denaturing gradient gel electrophoresis (PCR-DGGE) (T. W. Kim *et al*., 2010; Tanaka *et al*., 2012). Although these culture independent protocols provided insights into the general microorganism population during the fermentation process, it introduces bias in

determining microbial abundance (Hong *et al*., 2009; Nikolausz *et al*., 2005; Polz & Cavanaugh, 1998). On the contrary, analysis through whole-genome shotgun (WGS) sequencing provides useful data pertaining to not only the taxonomic perspective but also metabolic capabilities and functional diversity of the microbial communities during fermentative stages (Eisen, 2007; Gilbert & Dupont, 2011; C. Simon & Daniel, 2009). WGS therefore, provides comprehensive information into the soya sauce fermentation brine. Thus in this study, WGS approach using MiSeq (Illumina) platform and various bioinformatics tools were employed to obtain the composition of the microbial community and its functional profiles.

The quality of soya sauce is determined by its chemical properties (Gao *et al*., 2011; Yong & Wood, 1974). Alteration in the chemical features in the soya sauce brine causes selective environment for microbial growth (Hutkins, 2006; Y. S. Kim *et al*., 2011; Noda *et al*., 1980) besides giving an insight into the microbial carbon utilizations. The changes in chemical composition during the fermentation process were monitored by titration methods, gas chromatography (GC) as well as Kjeldal method.

The objectives of this research project include the following:

1. To monitor the changes in physicochemical properties and its relation to microbial growth during the traditional Chinese soya sauce brine fermentation process.
2. To identify the presence of bacterial and fungal population involved in traditional Chinese soya sauce brine fermentation process using 16S rRNA and 18S rRNA genes, respectively via next generation sequencing (NGS).
3. To observe the microbial succession during the fermentation process.
4. To identify the metabolic capabilities of the microbial population.

CHAPTER 2.0

LITERATURE REVIEW

2.1 Background of soya sauce

Humans have been fermenting food for thousands of years, evolving the process to achieve the required product. The ingredients used are generally sourced regionally, using materials that are readily available. Production of fermented food in Asia varies greatly from large scale factories such as those in Japan, China and Thailand, to medium and small scale manufacturing facilities found in Malaysia and most other Far East countries (Hutkins, 2006).

One fermented food that is ubiquitous in most Asian homes and food outlets is soya sauce. This important condiment is also known as *shoyu* (Japan), *kecap* (Indonesia), *jiangyu* (China), *Joseon ganjang* (Korea) and *kicap* (Malaysia). Besides being an essential ingredient in cooking and flavourings, soya sauce has been traditionally known to have medicinal properties such as easing the effects of menopause and reducing hypertension. It has been reported by Kinoshita *et al*. (1993) that soya beans, which is a key ingredient in soya sauce have high amounts of nicotianamine. Nicotianamine has been reported to suppress hypertension by inhibiting the angiotensin I-converting enzyme (ACE) (Hayashi & Kimoto, 2007). Soya sauce has also been described as a promoter in the secretion of gastric juice which promotes digestion (Kataoka, 2005). Investigations into soya sauce have also found it to have antimicrobial properties (Masuda *et al*., 2008), anti-platelet activity (Tsuchiya *et al*., 1999) and anti-allergic activities (Kobayashi, 2005). Therefore, soya sauce has been termed as a functional food besides being one of the core ingredients in Asian delicacies.

2.2 Microbial communities, chemical properties and soya sauce fermentation

Soya sauce is made through the fermentation of boiled soya beans, wheat grain or wheat flour, water and salt or brine. The fermentation process involves three major stages, namely, the *koji* production, brine fermentation and product refining (Luh, 1995).

Koji production is an essential starting material in most soya fermentation processes. *Koji* is defined as "mouldy grain" in Chinese (Hutkins, 2006). Also called soya bean *koji* or soya bean paste, it is made by soaking the soya beans for several hours before subjecting it to pressure cooking for an hour. Wheat and wheat flour is added into most soya bean *koji*, which enhances the flavour of the end product besides reducing the moisture content. The reduced moisture would lower the chances of undesirable bacteria from growing.

The second step involved in the production of soya sauce is the brine fermentation, where more than 20 % (w/v) of salt water is added into the prepared *koji*. This would further inhibit undesirable microorganisms due to the high salinity environment (Noda *et al*., 1980).

In the early stages of fermentation, the soya sauce mash would typically contain the starter strains, which in many cases would be *Aspergillus* species besides having the natural microbial community that exists in the *koji*. *Zygosaccharomyces rouxii* has also been reported to be present in the early stages of fermentation, as part of the natural *koji* flora (Tanaka *et al*., 2012). Some manufacturers also add yeast into the *koji* as a starter culture, which is known to produce the distinct peanut-like aroma (T. W. Kim *et al*., 2010; Zhao *et al*., 2009).

At the start of the fermentation process, lactic acid bacteria population is lower in abundance. During this stage, most of the microorganisms present are the *Bacillus* and *Micrococcus* species (Hutkins, 2006). However, Tanaka and colleagues (2012),

who used PCR-DGGE, reported that the *Bacillus* and *Micrococcus* species were not found in their study and concluded that these bacterial population are contaminants in the soya sauce *koji*. Instead, the above mentioned study showed presence of *Weissella cibaria, Lactobacillus fermentum, Staphylococcus gallinarium, Staphylococcus kloosii* and *Staphylococcus arlette.* Not surprisingly, these lactic acid bacteria (LAB) were also reported in studies using nested PCR-DGGE on *miso* (Japanese fermented soya beans paste) and Chinese fermented soya bean paste (Kim *et al*., 2010). Similarly, study using pyrosequencing noted that the population of *meju* (Korean fermented soya bean bricks) was also dominated by the LAB and *Bacillus* species (Kim *et al*., 2011). The same was reported for the fermentation process studies conducted on rice vinegar (Haruta *et al*., 2006) and *kimchii* (Jung *et al*., 2011*).* Thus, the LAB was noted as microbial populations that are essential in food fermentation processes.

Later in the fermentation process of soya sauce, the changes in the environment of the mash makes it extreme to some of the organisms that was present in the earlier stages. The salinity level would increase to around 18 % after a month of fermentation and this would inhibit the growth of organisms that are incapable of tolerating such an environment (Hutkins, 2006; Kim *et al*., 2011; Noda *et al*., 1980). The high salinity, selective environment, instead favours the proliferation of yeasts (Betts *et al*., 1999) and other halotolerant microorganism. Similarly, this was observed in a study conducted on Korean soya sauce by Lee *et al*. (2006).

As the fermentation continues, the reducing sugar and total nitrogen level will increase due to the proteolytic and amylolitic activity. However, it was noted by Gao *et al*. (2011) that these parameters will drop after approximately 60 days into the fermentation process due to its utilization by yeast and the LABs. The pH level will also decrease as the fermentation progresses. The increase of acidity content is due to the increase of metabolic by-products, namely the lactic acids and acetic acids produced by

salt tolerant LAB, *Pediococcus halophilus* and *Tetragenococcus halophilus* (Noda *et al*., 1982; Zhao *et al*., 2009). The production of acids contributes to the distinct flavour of the end product and also inhibits the generation of low acid tolerant microorganisms. This includes *Bacillus* and *Micrococcus* species besides preventing the growth of pathogenic bacteria such as *Escherichia coli* O157:H7 (Kataoka, 2005).

In the middle stage of fermentation, salt tolerant yeast such as *Candida versatilis* will be present (Tanaka *et al*., 2012). Both *Z. rouxii* and *C. versatilis* were found to be more salt and pH tolerant than their LAB counterparts (Betts *et al*., 1999; van der Sluis *et al*., 2001). During this stage of fermentation, acid content accumulates progressively and the pH value drops to as low as 5.0. Even the growth of the halotolerant LAB – *T. halophilus* and *Lactobacillus delbrueckii* will be inhibited and yeasts begins to dominate the mash population (Hutkins, 2006). However, it was reported by Zhao *et al*. (2009) that the presence of *T. halophilus* was not affected throughout the fermentation of Chinese soya bean paste.

At the end stage of soya sauce fermentation, further inhibition of microorganisms occurs as the pH value reaches to approximately 4.0 and a salt concentration of about 20 %. Tanaka *et al*. (2012) observed the presence of halotolerant *Candida etchellsii* by the last stage of fermentation of *shoyu*. It was stated by Hutkins (2006) that by the end of soya sauce fermentation, even the proliferation of *Z. rouxii* would be inhibited. Interestingly, it was noted by Noda *et al*. (1980) that the food spoilage yeast *Saccharomyces rouxii* in some cases was able to survive the extreme environment. It was observed in another study on food spoilage yeasts that pH and salt level has relatively high influence on the growth of yeasts (Betts *et al*., 1999).

2.3 Food bacteria culturability

The turning point in microbiology was the development of the culture media by Robert Koch in late 1800s and his hypothesis on the relationship of causative microorganism and diseases. His postulates require a microorganism to be cultured to cause an infection. Ultimately, this split the microbiological world into two "cult", namely the culturable and the unculturable (Handelsman, 2004). Scientists continued to argue that not all microorganisms are culturable as the number of cultured microorganisms does not equal to the number that was observed under the microscope. This is distinctively observed in aquatic environments (Kennedy *et al*., 2010) and the soil environment where only 1 % of the populations were reported as culturable (Torsvik *et al*., 1990).

In food microbiology, enrichments were often done to isolate bacteria or fungus of interests as the microorganisms are believed to be distressed by the processes involved in the food production (Pettengill *et al*., 2012). The antagonism that exists among the different organisms in the same environment and the different rates of growth, develop enrichment biases when it comes to detecting pathogens and also in studying microbial populations. The faster growing organism would essentially thrive and thus, limit the diversity or growth of other microorganism (Dunbar *et al*., 1997). Recognizing the fallback of pure culture isolation methods in the study of microorganisms, scientists today turn to a new way towards understanding the microbial world through metagenomics.

2.4 Metagenomics

Metagenomics is an approach which enables scientists to study uncultured microorganisms. Metagenomics offers a continuous opportunity for scientist to tap into microbial communities of extreme environments, without the limitations of culturing.

Metagenomics also offers an unlimited tool to further understand the functionality of the microbes. It was noted by Gill *et al*. (2006) that the genomes of the microbial communities in the human gut exceeded those of the human genome. This proved that the microbial community contributes to the health and diseases within the human body. In addition, metagenomic approaches allow us to develop a better system in bioremediation, energy and in combating diseases (Handelsman *et al*., 2007). Metagenomic also provides the opportunities to even manipulate the micro-ecosystem, enabling us to improve food processes and the final food products.

Metagenomics is characterized into two different approaches (Figure 2.1), namely the functional based metagenomics and the sequence based metagenomics (Handelsman *et al*., 2007). The functional metagenomic analysis involves DNA cloning and screening these DNA for the traits of interest, while the sequence based metagenomic analysis looks into massive amounts of sequences which will be screened for targeted traits of interest (Handelsman, 2004; Simon & Daniel, 2011).

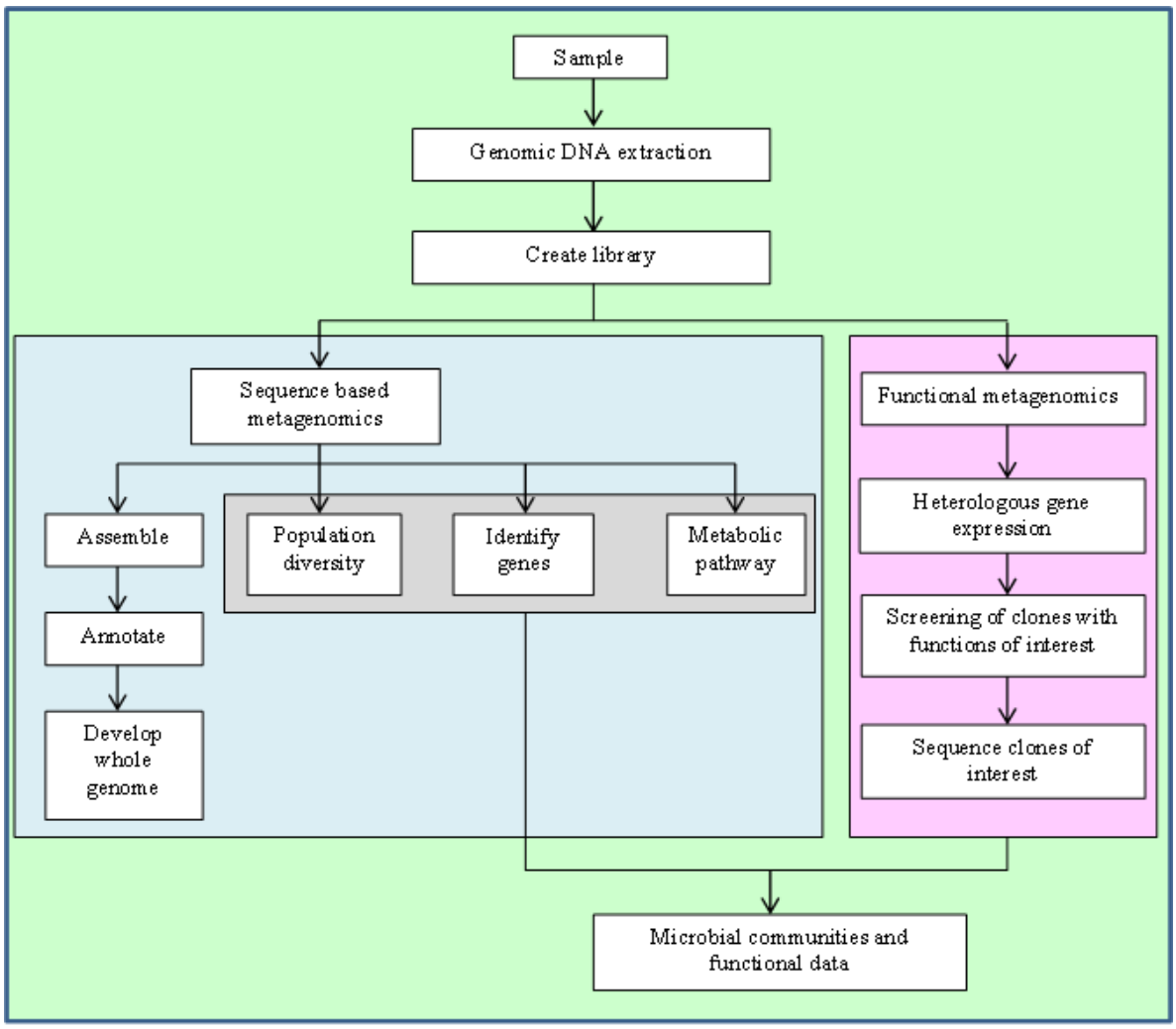

Figure 2.1 Functional and sequence based metagenomics. Adapted from Handelsman *et al*. (2007). The functional based metagenomics requires screening through thousands of clones for the character of interest. The sequence based metagenomics involves more extensive sequence analysis with known databases.

2.4.1 Functional metagenomics

Functional metagenomics as its name suggests, involves screening of DNA library clones for their phenotypic traits of interest. This approach requires the libraries to be prepared by transforming the DNA extracted into a cloning vector. These metagenomic libraries or clones are then subjected to functional screening. Metagenomic studies using this approach have successfully identified novel genes. A study by Gillespie *et al*. (2002) successfully isolated *turbomycin A* and *turbomycin B*

from soil DNA. These compounds presented antibiotic property against both Gram positive and Gram negative bacteria. Another study using cloned libraries from wastewater plant was able to identify extradiol dioxygenases (EDOs) genes which is important in the degradation of aromatic compounds in the environment (Suenaga *et al*., 2007).

The approach of functional metagenomics does not require identification by sequence analysis, thus does not require comparison with a known sequence databases. The benefit of this extremely selective method is that a novel functional gene could be identified precisely (Simon & Daniel, 2011). Findings using this approach produce very few false conclusions, unlike the sequenced based metagenomics where false annotation could be produced due to low similarities in the reads (Kennedy *et al*., 2010).

However, functional metagenomics has its disadvantages. It was noted by several studies that this approach introduces bias as not all genes can be expressed in a cloning vector (Daniel, 2005; Handelsman, 2004).

2.4.2 Sequence based metagenomics

Sequence based metagenomics is an approach where primers are designed to target highly conserved region of known genes or phylogenetic targets (Simon & Daniel, 2011). Sequence based metagenomics approach has successfully been applied particularly in isolating novel enzymes. For example, LeCleir *et al*., (2004) isolated novel chitinase sequences from aquatic samples. This discovery is important as chitinase is responsible for the degradation of biopolymer chitin which is crucial for the environment. Another discovery based on PCR screening was in 2008, when Banik and Brady found two new glycopeptide biosynthetic gene clusters which are important in the development of new antibiotics against pathogenic Gram-positive bacteria.

The sequence based metagenome approach has also been used in determining the microbial diversity in an environment. Conserved hyper-variable region, 16S rRNA gene assignments were used to study the microbial communities and to monitor the microbial succession trough time based in a sample of interest (Li *et al*., 2011). For eukaryotic microorganisms, namely fungus and yeasts, the highly conserved 18S rRNA gene or the internal transcribed spacer region 1 and 2 (ITS1 and ITS2) is targeted. In 2006, Haruta and friends reported the microbial succession and the major contributors of rice vinegar fermentation process by PCR-DGGE targeting both the 16S rRNA genes and 18S rRNA genes.

However, culture free protocols such as the PCR-DGGE and amplicon based metagenomic through targeted gene amplifications present partiality in establishing microbial distribution (Nikolausz *et al*., 2005; Polz & Cavanaugh, 1998). PCR-induced bias is shown in multiple magnitudes that would result in PCR template selection. Careful considerations and steps in the PCR methods have to be taken to minimize these errors (Acinas *et al*., 2005; Polz & Cavanaugh, 1998). A study conducted in 2009 revealed that PCR amplification of 16S rRNA genes from environmental samples will lose half of the bacterial population (Hong *et al*., 2009). Thus, whole genome shotgun sequencing (WGS) is an alternative to overcome the drawbacks of selective target amplifications (Simon & Daniel, 2009).

The shotgun approach in metagenomics were used in the determination of the Sargasso Sea microbiome, where over one billion base pairs of sequences were generated and over 1800 species of microorganisms were identified (Venter *et al*., 2004). Nevertheless, it was argued that the shotgun approach to the determination of microbial diversity does not allow in-depth detection of rare species in a complex community (Shah *et al*., 2011). Interestingly, studies by Kalyuzhnaya *et al*. (2008) on Lake Washington noted that both approach using the 16S rRNA gene amplifications

and shotgun metagenomic resulted in similar microbial distribution. In a separate study, it was noted that the random approach was a dependable method to identifying a microbial population (Manichanh *et al*., 2008).

2.5 Metagenomics and next generation sequencing (NGS)

Metagenomic analysis has opened up a whole new world into the study of microbial populations and has enabled in-depth knowledge of their presence and the discoveries of genes with promising functions. New technologies were introduced recently to facilitate in unraveling even further of the unknown microbial world. These new techniques in sequencing, coined as NGS are able to produce millions of reads, more economical and require shorter time for analysis.

Many metagenomic researchers have used the NGS in obtaining the microbial distribution. For example, Edwards *et al*. (2006) was able to compare the microbial population that exists in the Soudan mine using the Roche 454 Life Sciences pyrosequencing technologies with that of previously studied Iron Mountain communities using the Sanger sequencing method. In the medical field, NGS approach in metagenomics was able to determine the microbial population in the human gut and also determine the impact of long term antibiotic treatment on human intestinal microbiota (Jernberg *et al*., 2007). Hence this is important in understanding the relationships between the microbial community and human health (Gill *et al*., 2006). Not surprisingly, even studies on food technologies are not left behind in applying the NGS based metagenomics approach. For example, the Roche 454 FLX platform was used in the identification of the artisanal cheese and cheese rinds bacterial communities (Quigley *et al*., 2012). This is important in further understanding the factors that contributed to the different types of artisanal cheeses.

However, NGS as any scientific method has its limitations. The reads that are produced using this technology are short compared to the 800-1000 bp produced by Sanger sequencing (Shah *et al*., 2011). The lists of NGS platforms and their read lengths, as well as their strength and limitations are as stated in Table 2.1. The short reads and high throughput data place a challenge to decipher and maximize the value of the generated reads. Thus, to encounter this limitation, computational systems and softwares have been developed (Huson *et al*., 2009).

Table 2.1 List of NGS platforms, read lengths, expected throughput and error rates

Platform	Read length (bp)	Throughput/run (Mb)	Error rate %
Roche			
454 FLX+	700	900	1
454 FLX Titanium	400	500	1
454 GS	400	50	1
Illumina			
GAIIx	2 x 150	96,000	>0.1
HiSeq 2000	2 x 100	400,000	>0.1
HiSeq 2000 V3	2 x 150	<600,000	>0.1
MiSeq	2 x 150	1000	>0.1
Life Technologies			
SOLiD 4	50 x 35	71,000	>0.06
SOLiD 5500xl	75 x 35 PE 60 x 60 MP	155,000	>0.01
Ion torrent			
PGM 314 Chip	100	10	1
PGM 316 Chip	100 +	100	1
PGM 318 Chip	200	1000	1
Pacific biosciences			
RS	1500	45/SC	15

Adapted from Scholz *et al*. (2012).

CHAPTER 3.0

MATERIALS AND METHODS

3.1 Chemical reagents

All the chemical reagents used in this study are of analytical grade purchased from Bio-Rad Laboratories Ltd., U.S.A.; Promega Ltd, U.S.A.; Invitrogen Corp., U.S.A.; Epicentre, U.S.A. and Illumina, Inc., U.S.A.

3.2 Buffer solutions

Preparation of in-house solutions and buffers used in this study were sterilized by autoclaving at 121 ˚C, at 15 psi for 20 min. Heat sensitive solutions were filter sterilized with sterile 0.2 µm pore size nitrose cellulose membrane (Sartorius, Germany).

3.2.1 Phosphate buffered saline (PBS)

PBS buffer was prepared by weighing 0.23 g of NaH_2PO_4, 1.15 g of Na_2HPO_4 and 9.0 g of NaCl in 1 L of distilled H_2O. The pH of solution was then adjusted to the value of 7.0.

3.4.2 5× Tris borate EDTA (TBE) buffer

5× TBE stock solutions consisted of 54.0 g Tris base, 27.5 g boric acid and 3.72 g $Na_2EDTA.2H_2O$ dissolved in 1 L of distilled H_2O with pH adjusted to 8.0 before autoclaved. A 5× dilution was performed using distilled water to obtain 1× TBE.

3.4.3 Tris-HCl buffer (1 M, pH 8.0)

Tris-HCl buffer was prepared by mixing 121.1 g of Tris base with 800 ml dH_2O. The pH was then adjusted to the value of 8.0 with concentrated HCl. Distilled H_2O was added to 1 L. Solution was sterilized by autoclaving.

3.4.4 Ethylene-diamine-tetraacetic acid (EDTA) (0.5 M. pH 8.0)

EDTA (0.5 M) was prepared by mixing 186.1 g of $Na_2EDTA.2H_2O$ with 800 ml of distilled H_2O. The pH of solution was then adjusted to the value of 8.0, followed by sterilization by autoclaving.

3.4.5 1× TE buffer (10mM, pH 8.0)

A volume of 1 ml of Tris-HCl (1 M, pH 8.0) and 200 ml of EDTA (0.5 M, pH 8.0) was mixed with 98.8 ml of sterile distilled H_2O to make the 1× TE buffer (pH 8.0).

3.4.6 NaOH (0.2 N, pH 13)

A total of 0.4 g of NaOH was dissolved in 50 ml of ultrapure water. The pH was determined to ensure above value of 13.0. The solution was filter sterilized with sterile 0.2 µm pore size nitrose cellulose membrane. A fresh solution was prepared prior to each usage.

3.5 DNA marker

DNA ladder markers used in this study were GeneRuler^{TM} 1 kb DNA ladder purchased from Fermentas International Inc., Canada.

3.6 Sample collection

Chinese soya sauce fermentative samples were taken from a single tank with a volume of 145.0 L fermenting brine. The samples were taken from a traditional Chinese soya sauce factory in Malaysia which does not involve commercial starter culture. The soya sauce fermentation brine (3.0 L) was collected at day zero, 1, 2, 3, 4, 5 and 6 months of fermentation. Samples were collected in sterile bottles and stored at -20 ºC prior to DNA extraction and physicochemical analysis.

3.7 Physicochemical analysis

The reducing sugar levels were determined using the Lane-Eynon titration method. Total nitrogen and salt concentration was tested using the Kjeldal method and direct titration method based on Mohr method, respectively. Ethanol concentration was determined by gas chromatography (GC) and total acidity was measured based on direct titration method. Moisture content was measured based on oven method. All the above analysis except the pH determination was outsourced to accredited ISO 17025 certified laboratories.

3.8 Genomic DNA extraction

A volume of 2.5 L of soya sauce mash was centrifuged at 10,000 *g* for 45 min at 4 ºC. The pellet were washed 3 times with PBS (pH 7.0) and filtered through 4 layers of sterile gauze. The filtrate were sequentially filtered through nitrose cellulose membranes (1.2 µm and 0.2 µm pore sizes, respectively) using a vacuum pump. The DNA was harvested from the membranes and was extracted using the Metagenome DNA Isolation Kit for Water (Epicentre, USA) with minor modifications. The procedures are as stated below.

The membranes were cut and placed in a sterile tube with 1 µl of filter wash buffer and 0.2 % (v/v) Tween 20. The tubes were vortexed from low speed to the highest speed for 2 min with intermittent breaks. The cell suspensions were transferred into sterile microcentrifuge tubes and centrifuged at 14,000 *g* for 2 min. Supernatant was discarded and the pellets were resuspended with 1 ml 1× TE Buffer (pH 8.0). The cell suspensions were centrifuged at 14,000 *g* for another 2 min and the supernatant were discarded. The pellet was resuspended with 300 µl of 1× TE buffer (10 mM, pH 8.0), 2 µl of Ready-lyse Lysozyme Solution and 1 µl of RNase A. The cell suspensions were frozen at -20 ºC for 16-18 h (Kohn, 1960) before incubating at 37 ºC for 1 h. A volume of 300 µl of Meta-Lysis Solution (2x) and 1 µl of Proteinase K was added into the tubes. The tubes were incubated at 65 ºC for 15 min.

After incubation, the suspensions were cooled to room temperature for approximately 2 min prior to placing on ice for 5 min. Then, 350 µl of MPC Protein Precipitation Reagent were added and vortexed vigorously for 10 s. The suspensions were centrifuged for 10 min at 14,000 *g* at 4 ºC. The supernatant were transferred into 2 ml tubes and 570 µl of cold isopropanol was added. The tubes were inverted 25 times. DNA was pelleted by centrifugation at 4 ºC for 10 min at 14,000 *g*. Supernatant was discarded and 500 µl of 70 % (v/v) ethanol were added to the pellet. The tubes were centrifuged at 4 ºC for 10 min at 14,000 *g*. The ethanol was removed and the access ethanol was air dried for 15 min prior to the addition of 40 µl of 1× TE buffer (10 mM, pH 8.0) to elute the DNA.

3.9 Purification, quantification and quality control of genomic DNA

The extracted DNA was purified using the Agencourt AMPure XP (Beckman Coulter, USA) according to the manufacturer's protocol. DNA quality was checked on NanoDrop 2000c (Thermo Scientific, USA) and quantified on *Qubit* 2.0 (Life

Technologies, USA) fluorometer. The extracted DNA was amplified for 16S rRNA to ensure that inhibitors that could disable downstream work were not present.

3.10 Polymerase chain reaction (PCR) amplification

The PCR amplification was done on the extracted DNA before subjected to sequencing library preparation. The mixture is as stated in Table 3.3 below. The PCR cycles for the amplification of 16S rRNA region consisted of initial denaturation at 94 ºC for 5 min, followed by 30 cycles at 94 ºC for 30 s, annealing at 63 ºC for 30 s, extension at 72 ºC for 1 min 30 s and final extension step for 5 min. A negative control was included for each PCR run by substituting genomic DNA with ultrapure water.

Table 3.3 PCR mixtures for 16S rRNA gene amplification.

PCR Component (Stock concentration)	Stock Concentration	Working Concentration
Ultrapure water	N/A	N/A
10× Buffer containing $MgCl_2$ (2 mM) solution	10×	1×
dNTP	200 µm	0.2 mM
Taq polymerase	5 U/µl	1 U/µl
27F forward primer [5'-AGAGTTTGATCMTGGCTCAG-3']	10 µM	0.4 µM
1525R reverse primer [5'-AAGGAGGTGWTCCARCC-3']	10 µM	0.4 µM
Genomic DNA		2.5 ng/µl

3.11 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out by loading 4 µl of amplified DNA samples and 0.2 µg of 1 kb DNA marker in 1 % (w/v) ethidium bromide (EtBr) stained

agarose gel. Electrophoresis was done at 80 V for 45 min. Upon completion, the gel was visualized on UVP ultraviolet trans-illuminator.

3.12 DNA library preparation

Sequence libraries were generated using the Nextera DNA Sample Preparation Kits (Illumina, USA) according to the manufacturer's guidelines. The Nextera DNA sample preparation procedures combined fragmentation, end-polishing and adaptor ligation steps into one tagmentation reaction. The libraries were quantified using *Qubit* 2.0 fluorometer. The quality and the size distribution following the tagmentation process of the libraries were determined by using High Sensitivity DNA chips and DNA Reagents (Invitrogen, USA) on BioAnalyzer 2100 (Agilent, USA). The DNA concentrations were adjusted according to the formula below:

$$\text{Quantity of the libraries generated (ng/µl)} = \frac{3\ \text{nM}}{(\text{Library average size in bp}/500)}$$

The calculated DNA concentration of 2 nM was denatured with NaOH (0.2 N, pH 13.0) and diluted to a final concentration of 8 pM prior to being loaded into the MiSeq reagent cartridge (Illumina, USA).

3.13 NGS of soya sauce metagenomic DNA

Sample sheets were generated using the Illumina Experimental Management software version 1.3. A volume of 600 µl of the 8 pM denatured DNA libraries was loaded into the 300 cycles MiSeq reagent cartridge (Illumina, USA) and sequencing was performed in the MiSeq system for a 151 cycle paired-end run.

3.14 Sequence analysis of extracted metagenomic DNA

The generated sequences were downloaded from BaseSpace (Illumina, USA) account and the data obtained were processed and trimmed using CLC Genomic Workbench 5.1.1 with Q20. Sequences less than 50 bp were discarded. Minimum contig length was set at 400 bp. Assembled contigs with an average coverage of more than 10 times were identified for the presence of eukaryotic components by BLASTN with GenBank nucleotide (nt) database with e-value of $< 10^{-6}$. The contigs were distributed into prokaryote and eukaryotic origin.

3.15 Taxonomic assignments of soya sauce brine microbiome

The unassembled quality trimmed sequences were subjected to USEARCH v6.0 (Edgar, 2010) with the SSURef_111_NR alignment from SILVA (Pruesse *et al*., 2007) for both the 16S rRNA and 18S rRNA with e-value of $< 10^{-20}$. The data obtained in the sequence analysis were analyzed for taxonomic distribution using Metagenome Analyzer (MEGAN) version 4.70.4 (Huson *et al*., 2011). Identity of matches assignments were based on minimum of 97 % at genus level. Microbial abundance was calculated based on number of assigned reads.

In order to compare the overall microbial community in each of the soya sauce brine fermentation stages and the sampling completeness, the sequences from all of the samples were clustered into operational taxonomical units (OTUs) based on 3 % sequence dissimilarity with Uclust and Good's coverage, respectively using an open source software, Quantitative Insights into Microbial Ecology (QIIME) (Caporaso *et al*., 2010).

3.16 Metabolic reconstruction of soya sauce brine metagenome

The classified prokaryotic contigs were screened for microbial gene prediction using Prodigal V2_60 (Hyatt *et al*., 2012). Independently, the eukaryotic contigs were subjected to web server gene prediction software, AUGUSTUS (Stanke & Morgenstern, 2005). Predicted query protein coding regions were compared to reference protein sequences in NCBI non-redundant (nr) database by using the protein database search tool RAPSearch V2_60 (Ye *et al*., 2011). Metabolic potential analysis was done based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) classification scheme visualized on MEGAN version 4.70.4. Metabolic pathways present were filtered using MinPath (minimal set of pathways) parsimony analysis (Ye & Doak, 2009). Statistical analysis is based on the total number of reads categorized in the system's categories.

3.17 Core metagenome of traditional Chinese soya sauce brine

The unassembled quality trimmed reads from all the seven stages of fermentation were combined and assembled via *de novo assembly* to establish the core metagenome. The regulatory pathway and protein functions of traditional Chinese soya sauce brine fermentation process were generated using iPath2.0 (Yamada *et al*., 2011).

In order to obtain the abundance of a specific gene of interest, the trimmed reads from each fermentation stage were mapped against the predicted core metagenome with 90 % identity and 80 % of the aligned length. Functional capabilities were then visualized using the KEGG classification scheme. The gene of interest were inspected and the sequence descriptions were verified using Blast2GO (Conesa *et al*., 2005). The calculation for abundance was done based on the number of total reads against the concensus length.

3.18 Nucleotide sequences accession number

The soya sauce mash metagenomic raw reads were deposited in the NCBI Short Read Archive with accession number SRA064709 under study accession SRP017928.

CHAPTER 4.0

RESULTS

4.1 Physicochemical analysis of soya sauce samples

4.1.1 pH and total acidity

The pH value and percentage of total acidity was measured periodically at day zero and later monthly until the end of the fermentation process. The pH values were at 5.3 at day zero and decreased to pH value of 4.37 by the end of the sixth month fermentation process (Figure 4.1). The percentage of acidity increased from 0.15 % (w/w) at day zero to 0.53 % (w/w) at the end of the fermentation process.

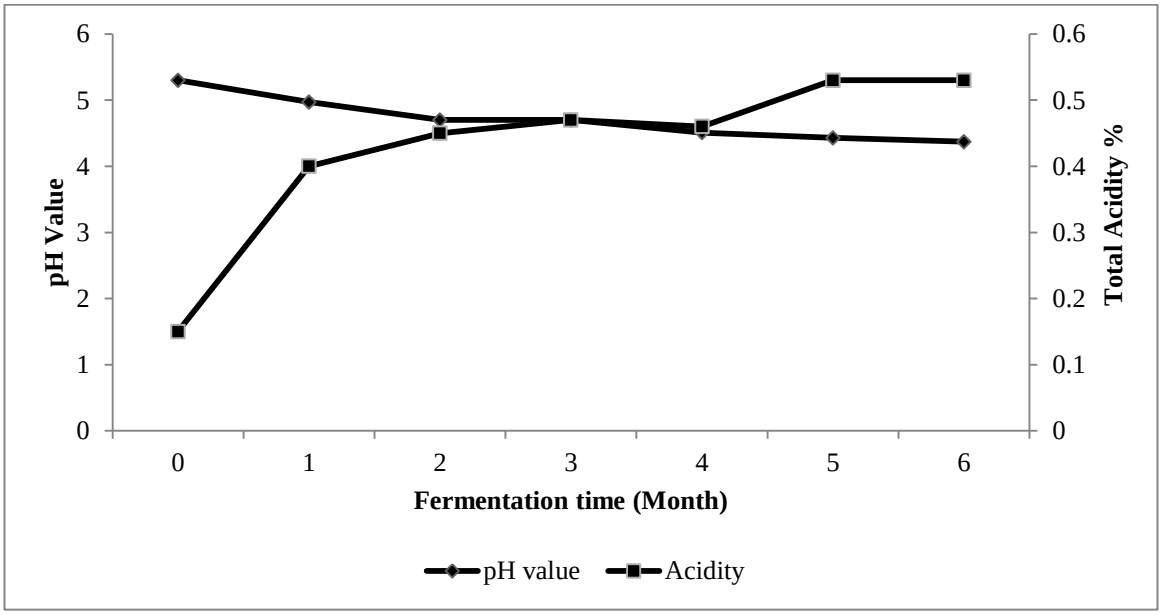

Figure 4.1 pH value and total acidity mean values at various fermentation stages. Data are expressed as the means of ± SEM value of triplicates (pH value) and duplicates (acidity) experiments. The pH decreases while the total acidity increases as the fermentation process progressed.

4.1.2 Salt concentration

The salt (NaCl) concentrations were measured at day zero and monthly until the end of the fermentation process. The soya sauce brine concentration was at 2.89 M at day zero and increased to as high as 3.43 M by the end on the fermentation process (Figure 4.2).

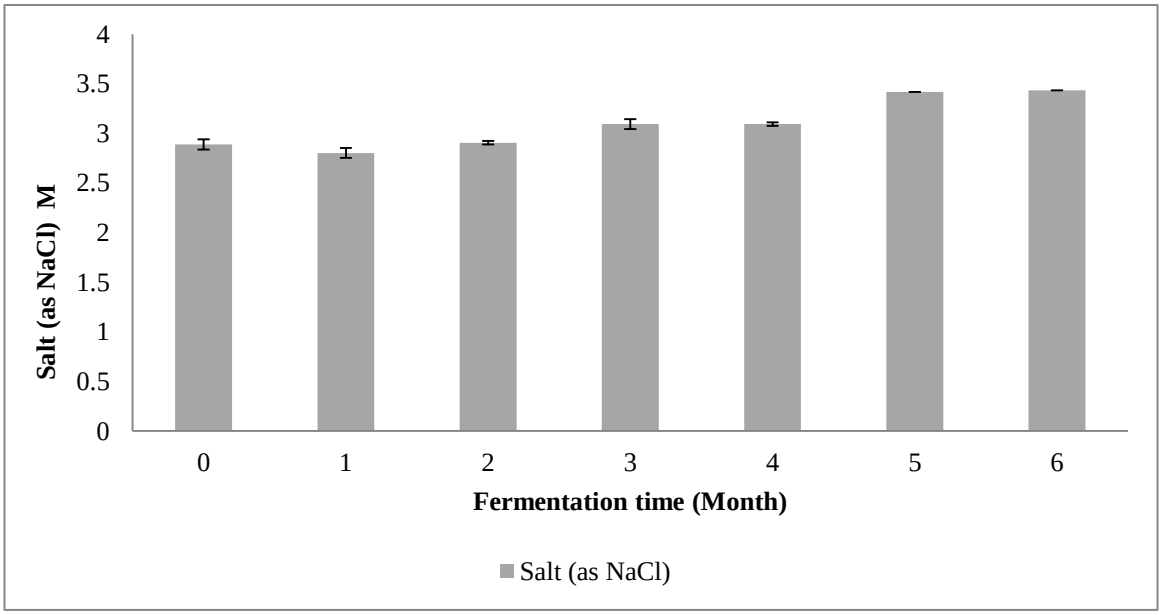

Figure 4.2 Salt content of the Chinese soya sauce mash. Data are presented as means of ± SEM values of duplicate analysis. The NaCl concentration increases as the fermentation process proceeded.

.

4.1.3 Ethanol concentration

The ethanol concentration was monitored at every sampling time. Initially, ethanol concentration was not detected (<10 mg/kg) at day zero and the amount increased as the fermentation progressed (Figure 4.3). By the end of the sixth months of the soya sauce fermentation, the ethanol concentration was at 1105.7 mg/kg, amounting to 0.1 % of the soya sauce brine. The reduction of 0.02 % from the fifth month at the end of the fermentation period is believed to have occurred through passive evaporation.

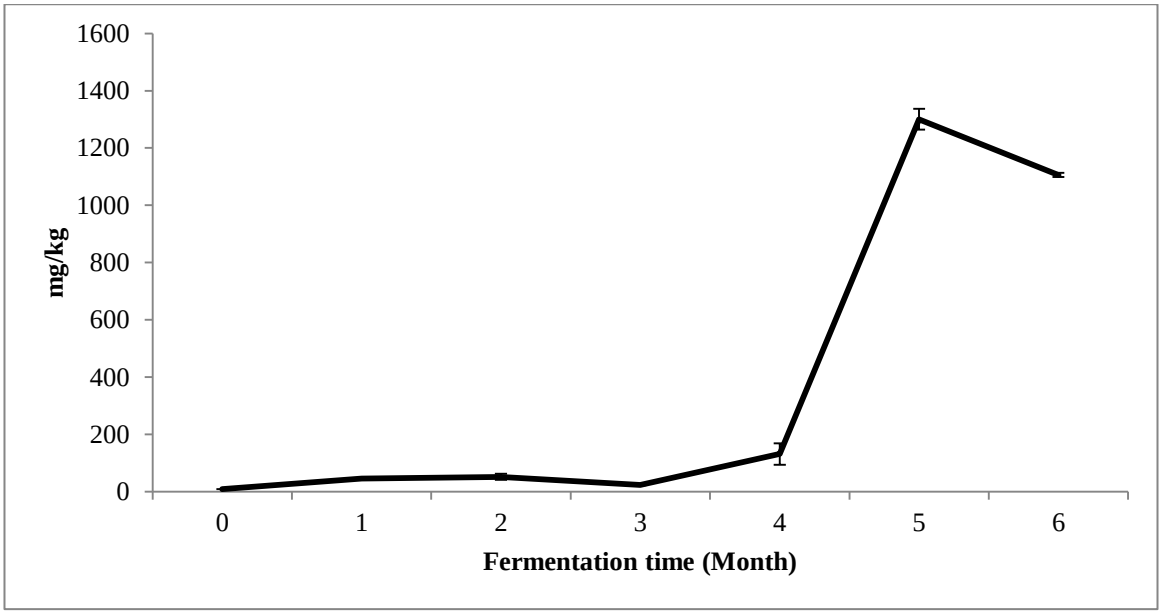

Figure 4.3 Changes in ethanol concentration of traditional Chinese soya sauce brine fermentation. Data are presented as means of ± SEM values of duplicate experiments.

4.1.4 Moisture and water activity (a_w)

Moisture content and water activity was tested on all the samples taken from the fermentation tank during the six months soya sauce fermentation process. The moisture content decreased from 74.2 % to 69.6 % as the fermentation progressed while the water activity was stable at an average of 0.75 throughout the fermentation process (Figure 4.4).

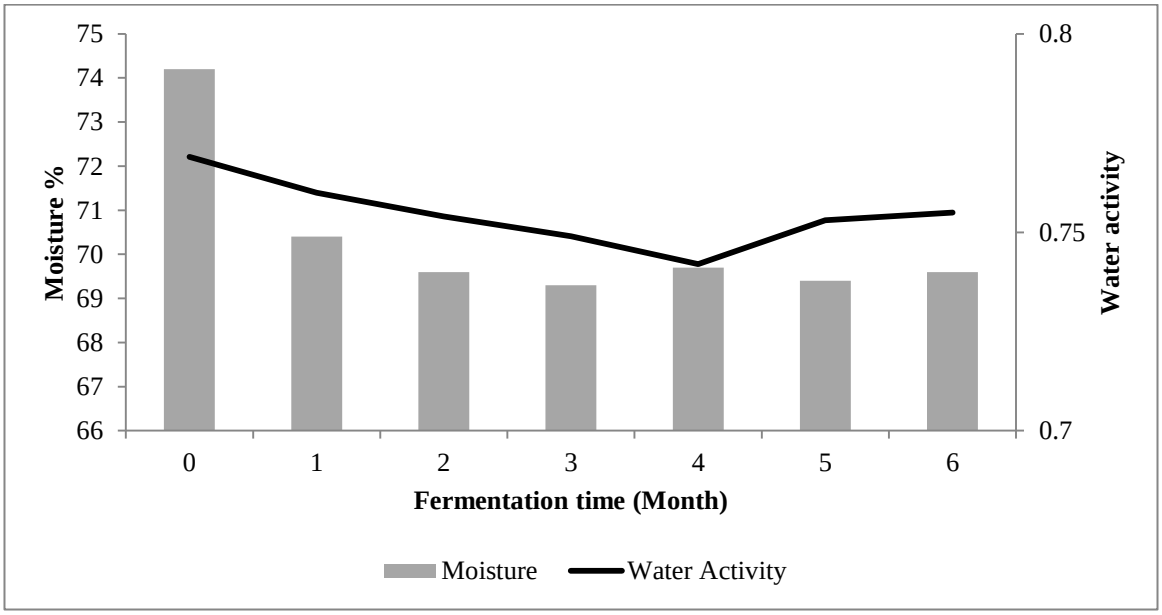

Figure 4.4 Changes of moisture and water activity. The moisture content decreases substantially after the first month while the water activity remains stable at an average of 0.75.

4.1.5 Reducing sugar and total nitrogen content

The reducing sugar and total nitrogen contents were investigated at each fermentation sampling stage. It was observed that the reducing sugar level increased in the first month and reached maximum level in the third month before decreased thereafter. By the sixth month of fermentation, the reducing sugar level dropped to less than 0.3 % (w/v). The total nitrogen level increased after a month into the fermentation process (Figure 4.5).

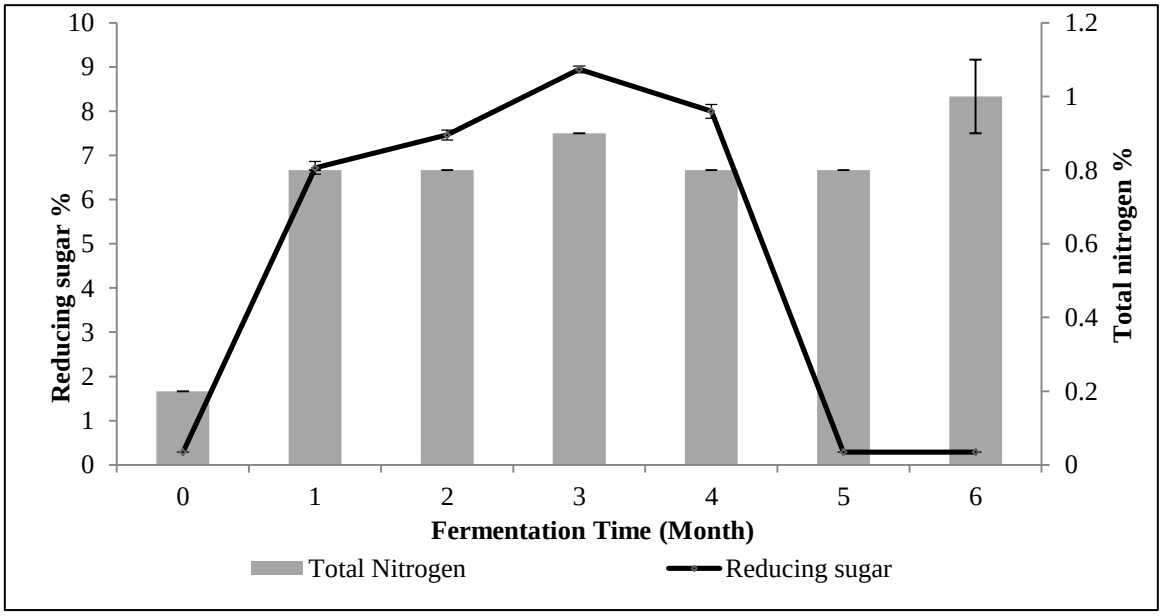

Figure 4.5 Changes in percentages of total nitrogen and reducing sugar content. Data are presented as average value of ± SEM values of duplicate experiments

4.2 Metagenomics DNA quantity and quality

The metagenomics DNA quality and quantity were checked prior to sequencing. Table 4.1 shows the results of the extracted traditional Chinese soya sauce brine genomic DNA.

Table 4.1: Quantity and quality of extracted genomic DNA

Fermentation time (Month)	$A_{260/280}$*	Quantity (ng/µl)
0	1.75	41.8
1	1.86	67.8
2	1.71	26.2
3	1.77	79.4
4	1.70	63.6
5	1.74	54.0
6	1.75	46.8

4.2.1 PCR amplification and inhibition determination

All extracted metagenomic DNA was amplified for 16S rRNA gene to ensure no inhibitory substance was present that could affect the downstream application. Figure 4.6 shows an example of 1.5 kb 16S rRNA amplification. The intactness of the extracted DNA was checked through agarose gel electrophoresis.

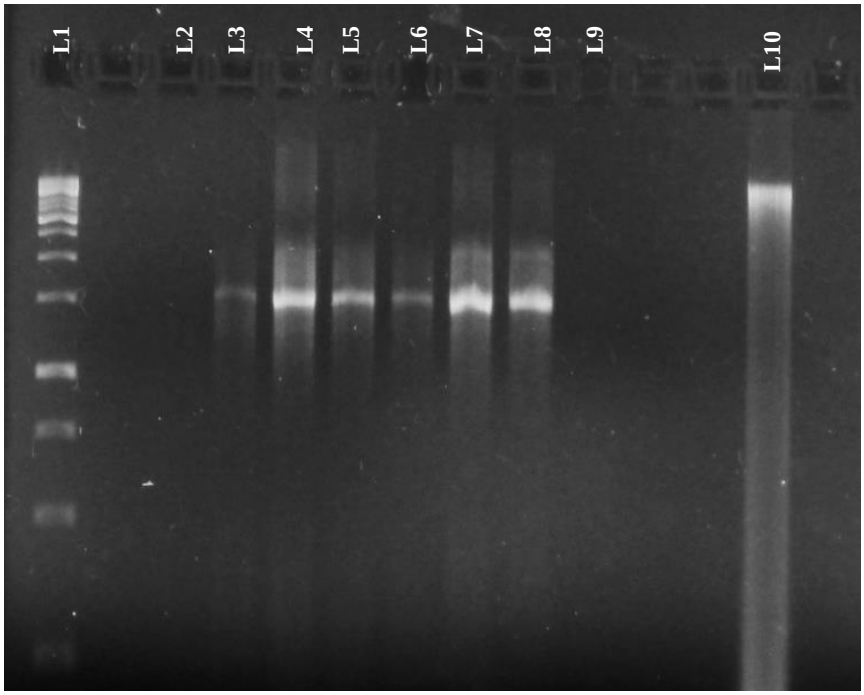

Figure 4.6 Agarose gel electrophoresis of 16S rRNA gene amplification from extracted soya sauce mash metagenomic DNA. L1: 1 kb DNA Ladder, L2: Negative control with ultra pure water as template. L3 – L9: Amplified 16S rRNA annealing temperature from 56 ºC to 63 ºC, L10: Extracted metagenomic DNA.

4.2.2 Library concentration

Prior to sequencing, the tagmented DNA is adjusted to the desired amount by using Qubit and calculation based on the size distribution on Agilent Technologies 2100 Bioanalyzer with a High Sensitivity DNA chip. This is to ensure that optimal cluster density can be obtained during sequencing. Figure 4.7 shows a typical libraries size distribution of a tagmented sample library. The size ranges from approximately 200 to 1000 bp, with the average size of 500 to 600 bp.

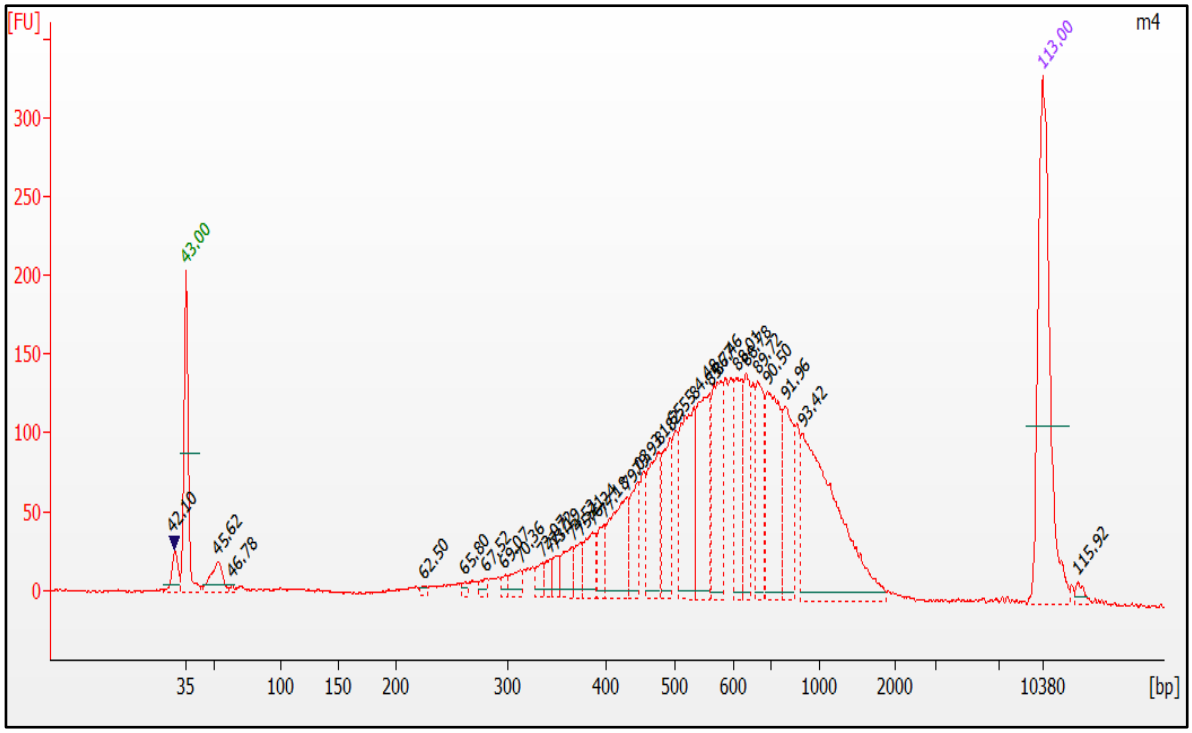

Figure 4.7 Determination of library size distribution. Tagmented sample library was analyzed for size distribution on the Bioanalyzer 2100.

4.3 NGS of traditional Chinese soya sauce brine fermentation

A total of 11.36 Gb of raw reads was generated from seven soya sauce brine fermentation sample. The quality reads after trimming was 9.10 Gb of sequences. An average of 1.30 Gb (ranging between 1.20 Gb to 1.96 Gb) of quality sequences were obtained from each sample. The average length of sequence generated after quality trim was between 112.2 bp to 134.2 bp. Table 4.2 summarizes the NGS data statistics from the soya sauce samples. The assembled sequences from all seven stages of soya sauce brine samples generated a total of 63,245 contigs of more than 400 bp in length, with N50 of 1,873 bp. More than 85 % of the genes had at least 50 % of their length covered by a single fermentation stage at 80 % identity threshold. This indicated that the genes of each fermentation stage includes more than half of the genes present in the fermentation tank.

Table 4.2: Summary of MiSeq sequencing data

	Month						
	0	**1**	**2**	**3**	**4**	**5**	**6**
No. of sequences generated (bp)	1,774,294,619	1,201,774,458	1,806,125,496	1,424,367,836	1,956,458,463	1,500,286,688	1,694,203,762
No. of reads generated	15,674,350	10,434,346	11,961,096	10,526,436	13,725,540	10,749,596	11,777,702
No. of quality sequences generated (bp)	1,505,234,972	1,078,183,875	1,182,280,801	1,210,597,794	1,582,353,972	1,213,475,521	1,316,079,598
No. of quality reads generated	13,976,473	9,482,454	10,820,621	9,630,743	12,420,866	9,545,509	10,550,831
Avg. length of quality reads (bp)	112.2	116.4	113.5	130.5	134.2	133.4	132.2
% of reads mapping to 16S rRNA[1]	0.67	1.0	0.77	0.55	0.25	0.38	0.23
% of reads mapping to 18S rRNA[1]	0.01	0.01	0.10	0.37	0.44	0.52	0.57
No. of contigs[2]	24,012	17,560	26,767	23,299	23,091	20,013	16,247
No. of prokaryotic genes predicted	17,704	11,679	28,501	12,786	10,981	8,708	6,867
No. of eukaryotic genes predicted	368	190	1,709	4,185	4,251	3,960	4,122
No. (%) of reads classified in KEGG	1,757 (9.72)	1,596 (13.45)	3,769 (12.48)	2,357 (13.89)	1,956 (12.84)	1,828 (14.43)	1,757 (15.99)

[1] Percentages calculated based on the total number of quality reads after removing sequences < 50bp.

[2] Calculated based on contigs of more than 400 bp in length

4.4 Microbiome of traditional Chinese soya sauce brine fermentation

An average of 0.55 % of the reads matched the 16S rRNA genes and 0.29 % of the reads matched the 18S rRNA genes at 97 % confidence threshold at genus level (Table 4.2). The analysis indicates the number of 16S rRNA genes decreased whereas 18S rRNA increased as the fermentation progressed. No archaeal 16S rRNA genes were found. The average number of OTUs for each sample was 1,360 which ranged between 1,040 OTUs to 1,800 OTUs. Rarefraction analysis to determine the sequence coverage of the current study showed similar pattern from each stages without reaching saturation (Figure 4.8). This suggests that a portion of OTUs still existed and that more sequencing attempt is required to detect all of the available phylotypes. However, based on Good's coverage in which estimates the sampling completeness, an average of 98.9 % (ranging in between 98 % to 99 %) with 97 % at genus level suggests that the majority of the microbial phylotypes were identified. As expected, like any approaches based on the SSU rRNA gene, gene copy number can confound abundance estimate (Kembel *et al*., 2012).

A microbial succession was observed over the six months fermentation process (Figure 4.9a). At day zero, *Weissella* (42 %) dominated the fermentation brine, followed by *Bacillus* and *Lactobacillus* with 22 % and 16 % respectively (Figure 4.9a). These three genera remained the most abundant population in the soya sauce brine until the third month and declined by the fourth month, to a total of 19 % by the end of sixth month. Throughout the fermentation process, lactic acid bacteria (*Weissella, Lactobacillus, Leuconostoc, Pediococcus, Enterococcus* and *Lactobacillus*) contributed more than 43 % of the population (Figure 4.9a).

The yeast community began to increase in the second month of fermentation. *Candida* spp. accounted for only 0.03 % of the population at day zero but by the sixth month, *Candida* population represented to more than 68 % of the soya sauce

microbiome. *Aspergillus* spp. maintained at around 0.1 % throughout the fermentation process. The dominancy changes between the yeast and LAB population is presented in Figure 4.9b. Overall, the biodiversity richness of bacterial population throughout the fermentation stages maintained throughout the process and increased in the final stage of fermentation. On the contrary, the diversity richness of fungal population displayed almost two fold-lower Shannon-Weaver index from 2.166 at day zero to 0.33 by the end of the fermentation process.

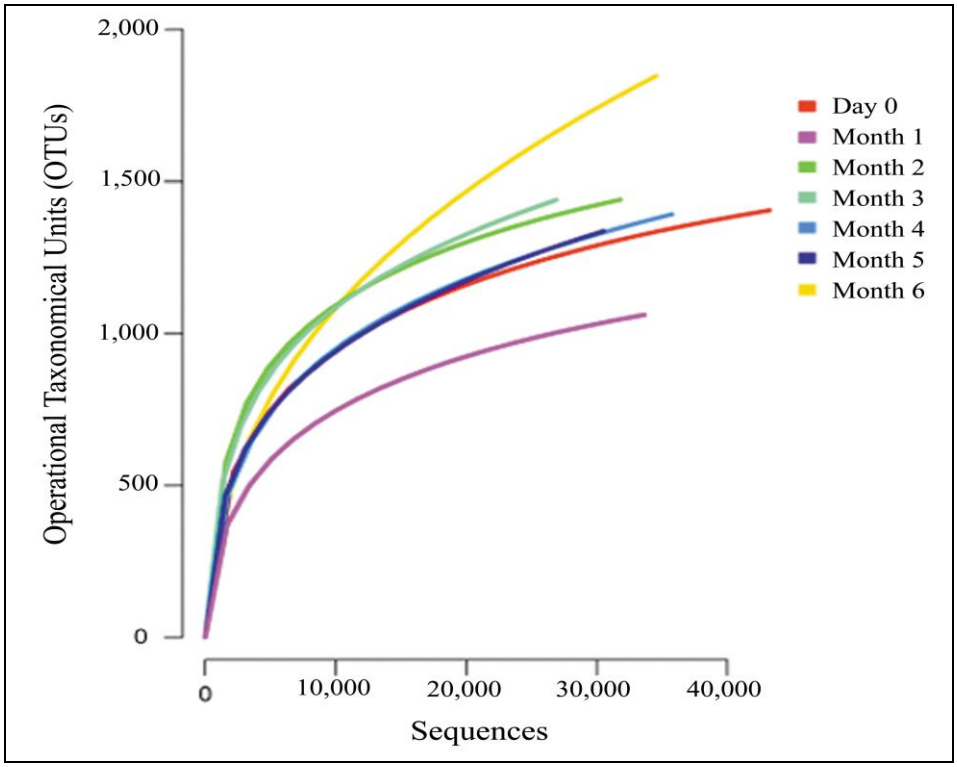

Figure 4.8 Rarefraction analysis of the traditional Chinese soya sauce fermentation brine microbiome. Rarefraction curves were constructed at genus level with 97 % sequence similarity.

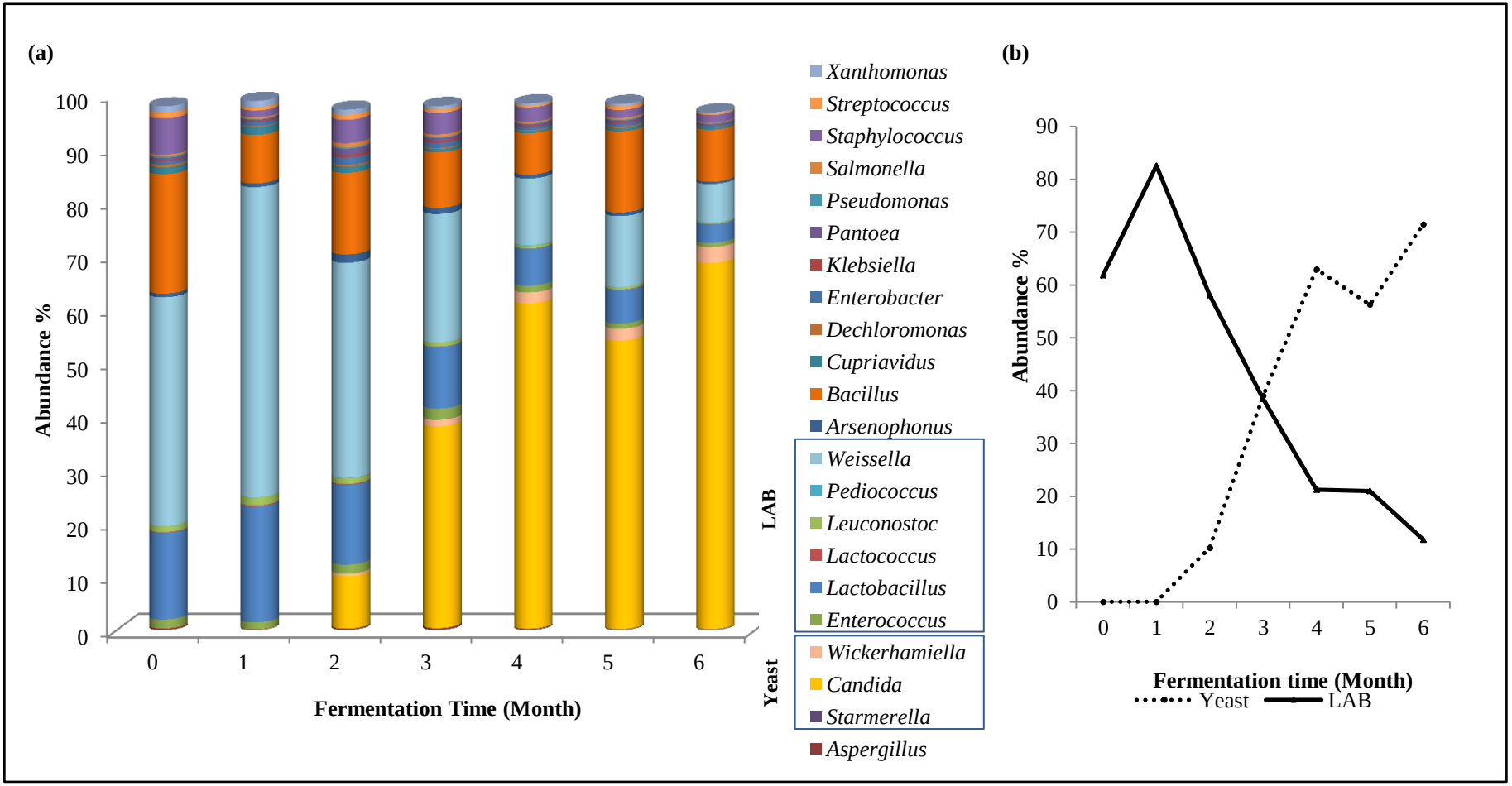

Figure 4.9 Microbial diversity of traditional Chinese soya sauce fermentation brine. (a) 16S rRNA and 18S rRNA gene sequences using MEGAN version 4.70.4 based on 97 % identity match at gene level and (b) Abundance of yeast and LAB in soya sauce mash.

4.5 Functional potential of traditional Chinese soya sauce microbial communities

The study identified a total of 1,799 orthologs matches the KEGG metabolic pathway nodes with 176 predicted pathways, providing a better understanding into the microbial functional capabilities during the fermentation period. An average of 13.26 % (ranging from 9.72 % to 15.99 %) genes was assigned to the KEGG pathways (Table 4.2). Metabolic reconstruction using KEGG showed that an average of 3,872 coding sequences (CDS) (16.96 %) was classified under carbohydrate metabolism and was consistently found in all of the soya sauce metagenome libraries, followed by amino acid metabolism, nucleotide metabolism and energy metabolism with an average of 2,404 CDS (10.53 %), 1,666 CDS (7.29 %) and 1,441 CDS (6.31 %), respectively (Figure 4.10).

Not surprisingly, the annotated functions are important metabolic pathways for all bacteria, which includes glycolysis (Figure 4.11a), pentose phosphate (Figure 4.11b), citrate cycle, protein complexes (oxidative phosphorylation) and cellular metabolism (purine and pyrimidine). Third level analysis of KEGG pathway found reputed fermentation specific functions which include the production of flavouring and aromatic compounds (arginine, proline, alanine, aspartate, glutamate). The collected data also showed evidence of potential methane metabolism. Further examination into the orthologous gene families leads to the findings of three unique functions to the pathway, namely EC 4.4.1.22 (S-(*hyroxymethyl*) glutathione synthase), EC 1.1.1.284 (S-(*hydroxymethyl*) glutathione dehydrogenase) and EC 3.1.2.12 (S-*formylglutathione* hydrolase). These enzymes are important in the detoxification of formaldehyde in bacteria (Goenrich *et al*., 2002).

Inspection into the metabolic potential of traditional Chinese soya sauce revealed the microbial capability for adaptation, whereby the genes involved in

oxidative phosphorylation revealed the presence of cytochrome *bd* complex. This enables enhanced microbial affinity to oxygen and is able express under limited oxygen condition (Borisov *et al*., 2007). Furthermore, gene encoding for subunit H of the vacuolar H^+-ATPase was absence in the microbial population of traditional Chinese soya sauce. Inspection of the glycolysis pathway revealed that the abundance of gene encoding for the production of the enzymes L-lactate (EC 1.1.27) and D-lactate (EC 1.1.1.28) dehydrogenase decreased over time. However, the fraction of gene covered for branched-chain amino acid aminotransferase (EC 2.1.6.42) in valine, leucine and isoleucine biosynthesis and degradation increased steadily after the first month (Figure 4.12).

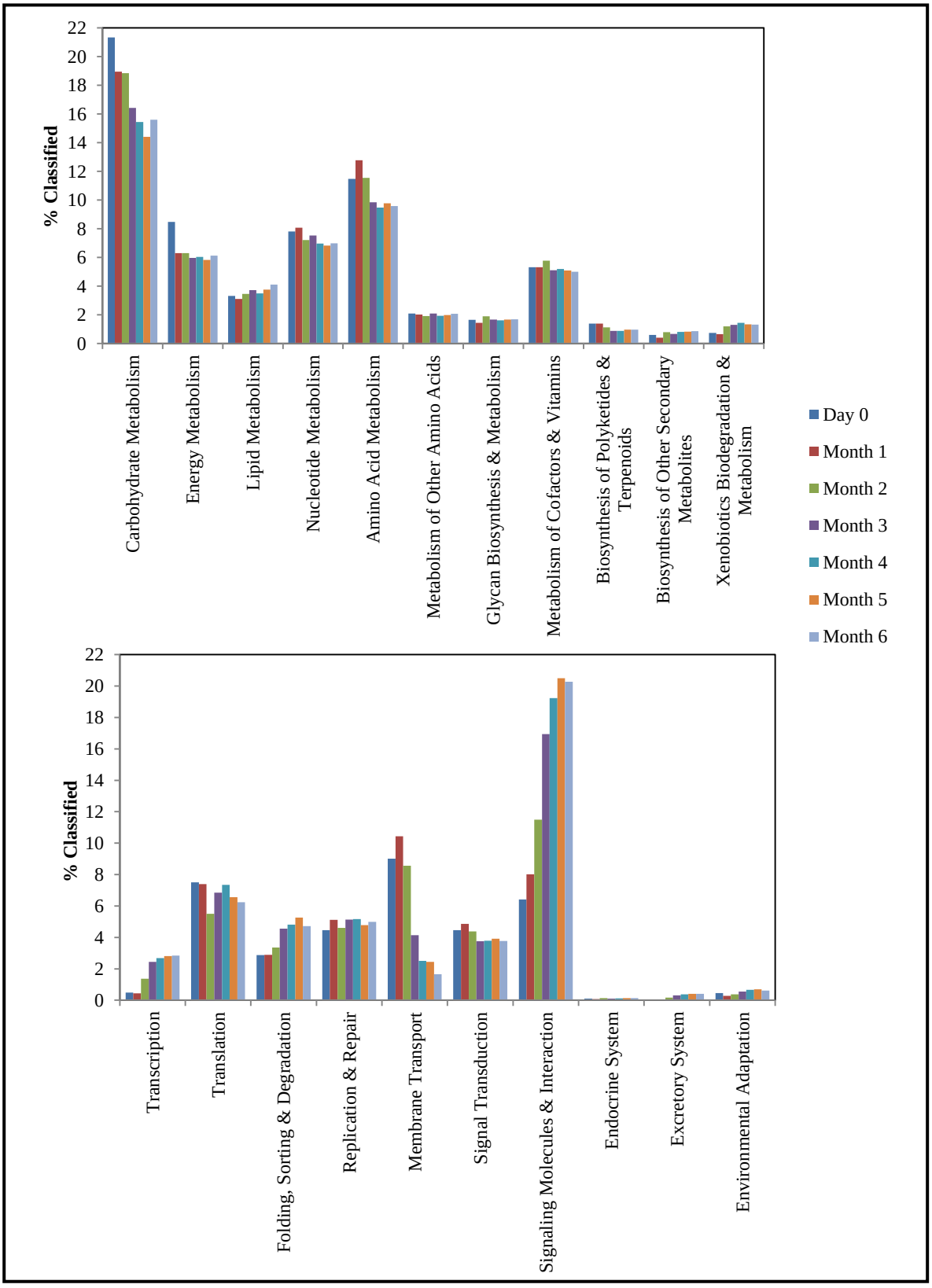

Figure 4.10 Traditional Chinese soya sauce fermentation brine functional capabilities. Predicted functional categories expanded to the second level of the KEGG classification.

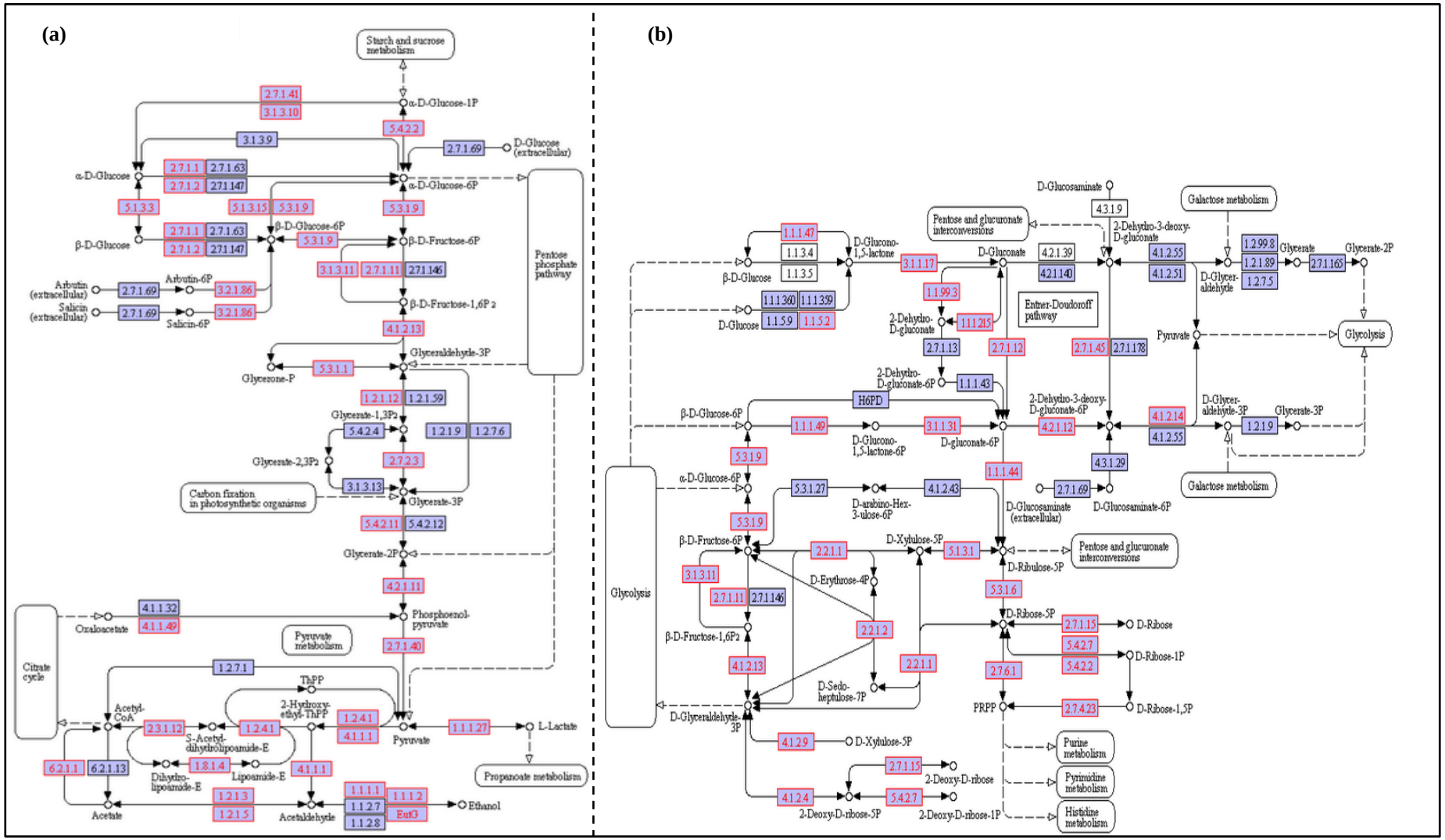

Figure 4.11 Presentation of the KEGG pathways (a) Glycolysis and (b) Pentose phosphate pathway using MEGAN. Blue boxes are functions identified in KEGG database. Highlighted in red borders are the functions that are found in the fermentation tank.

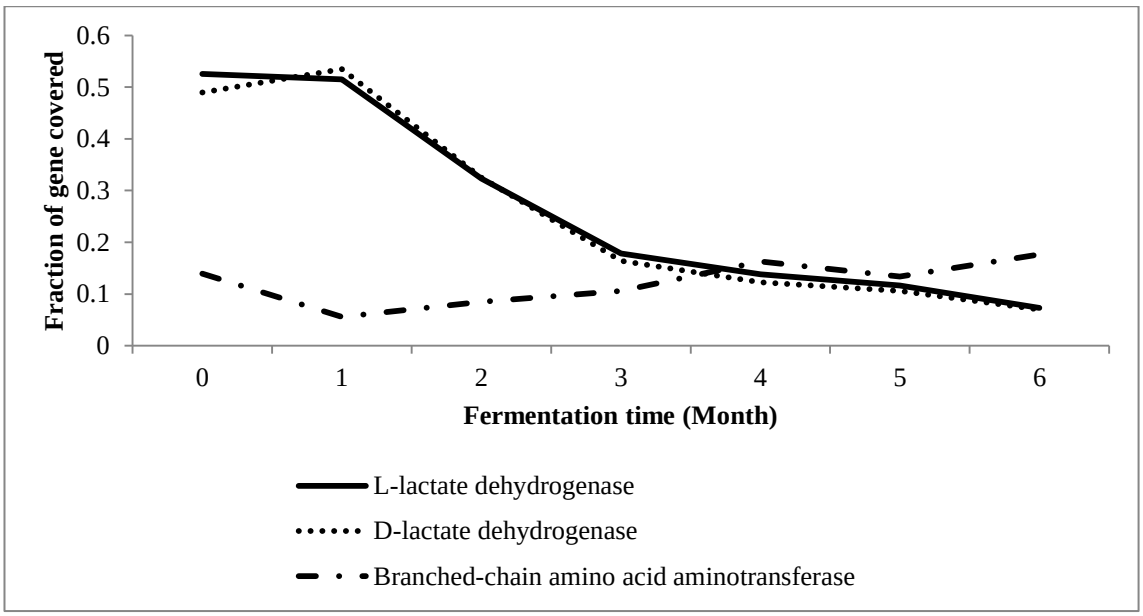

Figure 4.12 Coverage of gene encoding for L-lactate dehydrogenase, D-lactate dehydrogenase and branched-chain amino acid aminotransferase. The genes encoding for the targeted enzymes were based on KEGG pathway and verified with Blast2GO.

CHAPTER 5.0

DISCUSSION

WGS is a valuable molecular tool to study the microbial dynamics and functionality in a given natural environmental habitat by-passing the limitation and biases of culturing (Dunbar *et al*., 1997; Pettengill *et al*., 2012). In this study WGS metagenomic via NGS approach was used to establish the natural microbiota succession and their functional capabilities involved in the traditional Chinese soya sauce brine fermentation in addition to determining its relation to physicochemical changes.

5.1 Age-old soya sauce fermentation brine microbiome

Our NGS data revealed a clear succession of microbial assemblage in the traditional Chinese soya sauce fermentation brine. In the first half of the soya sauce brine fermentation, bacteria were the dominant microorganisms but succeeded subsequently by the yeast population. The genus *Weissella* represents the most abundant genus in the first half of the fermentation process and as the soya sauce aged, the yeast group *viz. Candida* genus succeeded to become the more dominant genus. Notably, albeit a minority, food-borne pathogens from the genera *Salmonella* and *Listeria* (Farber & Peterkin, 1991; Mead *et al*., 1999) were present in the soya sauce mash. As these bacteria represent only a small portion of the microbial population in the soya sauce fermentation brine, it was considered as contaminant from the environment most probably by the materials handlers. Their inability to survive in the fermentation tank was probably due to the stressful environment during the fermentation process and the production of antimicrobial peptides and proteins by the LAB (Ross *et al*., 2002; Rattanachaikunsopon & Phumkhachorn, 2010).

The gradual decrease of pH values and steady increase of acidity content over the six-month period of soya sauce fermentation, correlated to the presence of LAB. Our finding on the abundance of LAB present in our soya sauce samples is in agreement with previous study that has reported the isolation of LAB from soya sauce mash (Tanasupawat *et al*., 2002). It has been well documented that these salt tolerant LAB are responsible for breaking down carbohydrate into lactic acid and simple sugar, which caused acidification of the soya sauce mash (Kandler, 1983; Leroy & De Vuyst, 2004; Zhao *et al*., 2009). The decrease of pH level in the fermentative brine produced a hostile environment for less acid tolerant microorganism. In addition, the low water activity further inhibits undesirable microorganism from proliferating (Chirife & Buera, 1994). The pH value in the current study was lower than other reported study on soya sauce mash (O'toole, 1997; Yong & Wood, 1974). However, similar low pH value was recorded by a study on Korean soya sauce fermented in earthenware (Lee *et al*., 2006). Thus, we believe the lower pH reading was highly associated with geographical conditions and the raw material origins.

In addition to LAB, during the fermentation process of the soya sauce brine, yeasts, *Micrococcus*, *Streptococcus*, *Bacillus* and other related bacteria have been reported to be present in the soya sauce mash (Fukushima, 2004; Yokotsuka & Sasaki, 1997). However, it is noteworthy to point out that these published works are based on culturable bacteria study instead of using NGS as in the present work which will cover the unculturable microorganisms.

The increase in the brine salinity and acidity as the fermentation progressed provided an optimal growth environment for halotolerant and acid tolerant yeast (Betts *et al*., 1999; van der Sluis *et al*., 2001). The present study observed an increase of the yeast population by more than 70 % in which primarily involved the genera *Candida*, *Starmerella* and *Wickerhamiella*, whilst there was a reduction of LAB abundance by

50 % by the end of the fermentation process. Although the yeast population increased, the overall richness of the fungal diversity decreased over time. Even though the abundance of bacteria declined as the fermentation process proceeded, the bacterial diversity richness increased in the final stage of fermentation. This is due to the emergence of acid tolerant bacteria from the family *Propionibacteriaceae* (Jan *et al*., 2000) and *Acidobacteriaceae* (Hamberger *et al*., 2008). This correlates with the high acidic environment of the soya sauce fermentation brine by the end of the process. However, their existence only in the final month of fermentation may proof that they do not play an active role in the fermentation process. Further investigations are required to determine this. It is worthy to note that *Propionibacterium* has been used in the dairy industry as a secondary microflora to produce aromatic compounds and carbon dioxide (Giraffa, 2004).

The antagonistic environment in the fermentative brine was perhaps due to the chemical changes in selectively encourage the proliferation of certain group of microorganism while inhibits the other.

At the end of the fermentation process, we observe the presence of yeast genus *Saturnispora*. This yeast was reported to have been isolated from a broad array of environment. This includes estuarine water from mangrove forest, flowers, forest soil, insect frass, marsh water, rhizosphere of oyster grass and even *Drosophila* flies. This yeast has never before been associated with soy sauce fermentation before, although it has been isolated from Sauerkraut (Kurtzman *et al*., 2011). However, their function in biotechnology is unknown and raises a possibility of it being a potential spoilage microorganism (Kurtzman *et al*., 2011). Thus, our results suggest that extending fermentation time may encourage growth of spoilage microbial population.

5.2 Functional capabilities of the microbial assemblage

Pathway reconstruction is important to understand the biological functions involved in the fermentation process. However, pathway reconstruction posed a challenge. There exists redundancy in the estimation of pathways for a number of reasons. Firstly, different pathways can participate in the same biological functions (Okuda *et al*., 2008) and secondly, some proteins may carry out several biological functions (Rosin *et al*., 2005). Moreover, products from pathways or intermediates may be exogenously supplied. Finally, a protein may align to multiple homologous proteins with different biological functions through homology-based protein search (Ye & Doak, 2009). Thus, MinPath was utilized in this study. This web based software uses the minimum parsimony approach to eliminate spurious pathways that was over-estimated through naïve mapping approach (Ye & Doak, 2009). Therefore, this provided a more reliable estimation of pathways presented in the metagenome and consequently, the functional potential of the microbial consortia in the fermentation tank.

The data provided in this study were able to reveal the functional potential of the traditional Chinese soya sauce fermentation brine microbial community. The LAB found in the present study can be divided into two groups established according to the end product of glucose fermentation. *Lactococcus*, *Enterococcus* and *Pediococcus* are homofermantatives that produces sole lactic acid as the final product through Embden-Meyerhof-Parnas (EMP) pathway (Rattanachaikunsopon & Phumkhachorn, 2010). This explains the presence of genes encoded for the production of phosphofructokinase, pyruvate kinase and lactate dehydrogenase enzymes, which are crucial in the EMP pathway. *Weissella*, *Leuconostoc* and some of the lactobacilli are obligate heterofermentative LAB which produces lactate, carbon dioxide and ethanol and/ or acetate as a result of hexose monophosphate or pentose pathway (Battcock, 1998; König & Fröhlich, 2009; Wright, 2004).

We observed increased presence of gene encoding for the production of glucoamylase (EC 3.2.1.3) in the first month in which contributed to the increase of reducing sugar content in the first month as the enzyme hydrolyzed dextrin into simple sugars (Borglum, 1980). Meanwhile, the decrease in reducing sugar content in the fourth month onwards may be attributed to the rising yeast population which utilizes the simple sugars as carbon source. Similar findings have also been observed by Chou and Ling (1998). Additionally, the observed gene abundance for both D-lactate and L-lactate dehydrogenases decreased as the fermentation process progressed, exhibits the shift to heterolactic fermentation (Moat *et al*., 2003) in the fermentative brine. This confirms that traditional Chinese soya sauce in the present study is achieved mostly by heterofermentative LAB.

The emergence of yeast group in our samples correlated to the detection of ethanol in the soya sauce samples. The yeast population began to utilize simple sugars, producing ethanol as by-products. This finding is in agreement with other reported work observed by Ghose and Bandyopadhyay (1980). The amount of ethanol of 0.1 % (w/w) in the fermentation brine by the end of the process was typical of Chinese soya sauce (O'toole, 1997). However, the absence of the *Zygosaccharomyces*, and *Saccharomyces* which commonly attributed to the production of ethanol in fermented product (Hamada *et al*., 1990; Nissen *et al*., 2000), explains the lower amount of ethanol in the soya sauce brine in this study than that of the popular Japanese *koikuchi shoyu* (Luh, 1995; O'toole, 1997).

The presence of genes encoding for the branched-chain amino acids - isoleucine, leucine and valine (BCAAs) were observed from the soya sauce mash. BCAAs were reported to be important in some LAB for proteolysis (Liu *et al*., 2008) and essential in the production of volatile compounds such as acids, alcohol and esters (Marilley & Casey, 2004; Smit *et al*., 2005; Yvon & Rijnen, 2001). *Candida* species are able to

metabolize the BCAAs through the Ehrlich pathway which involves the branched-chain aminotransferase, decarboxylase, and alcohol dehydrogenase in producing flavour compounds in fermented food products (Derrick & Large, 1993; ter Schure *et al*., 1998). The gene encoding for the production BCAA aminotransferase increased as the fermentation process progressed. This corresponds to the increased abundance of *Candida* in our sample, which was found to be the most abundant microorganism by the end of the fermentation process. Interestingly, *Candida* was reported as producing higher amount of flavouring substances than any other yeast (O'toole, 1997).

The present study provides evidence of adaptive potential, especially in relation to oxidative phosphorylation. The presence of cytochrome *bd* complex coupled with high level of coverage in pentose phosphate pathway indicated reaction towards low oxygen degree in the fermentation tank. Furthermore, there was an absence of gene encoding for subunit H of the vacuolar H^+-ATPase. Subunit H is not required for the assembly of V-ATPase but an ATPase devoid of it cannot function and are unable to pump protons into the cell (Alba-Lois & Segal-Kischinevzky, 2010; Nishi & Forgac, 2002).

It was reported that lacking V-ATPase activity in yeast is unable to grow at pH neutral media, but thrive under a more acidic environment (Graham *et al*., 2003). This correlates back to our study, with the acidic condition in our soya sauce mash.

CHAPTER 6.0

CONCLUSION

Our meta-data showed the pH was closely associated with the abundance of LAB detected. The production of ethanol corresponded well to the shift to heterotrophic fermentation and the depletion of reducing sugar, in the spontaneous presence of yeast population. In addition to that, the total nitrogen liberated from soya and wheat flour protein increased without the addition of exogenous supply of nitrogenous sources.

The extensive biodiversity found in this study exhibits the complexities involved in this “ancient biotechnology” fermentation practice. The study was able to show the progression of both bacterial and fungal population throughout the fermentation process via WGS metagenomic coupled with NGS approach. Additionally, we were able to identify metabolic capabilities and functional diversity of the microbial communities during the fermentative stages. However, further metatranscriptomic studies are required to support our findings in regards to these functional properties.

REFERENCES

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology, 71*(12), 8966-8969.

Alba-Lois, L., & Segal-Kischinevzky, C. (2010). Beer & Wine Makers. *Nature Education, 3*(9).

Banik, J. J., & Brady, S. F. (2008). Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proceedings of the National Academy of Sciences, 105*(45), 17273-17277.

Battcock, M. (1998). *Fermented fruits and vegetables: A global perspective* (Vol. 134). Rome: Food & Agriculture Organization.

Betts, G., Linton, P., & Betteridge, R. (1999). Food spoilage yeasts: Effects of pH, NaCl and temperature on growth. *Food Control, 10*(1), 27-33.

Borglum, G. (1980). Starch hydrolysis for ethanol production. *American Chemical Society, Division of Fuel Chemistry, 25*(CONF-800814-P3).

Borisov, V. B., Forte, E., Sarti, P., Brunori, M., Konstantinov, A. A., & Giuffre, A. (2007). Redox control of fast ligand dissociation from *Escherichia coli* cytochrome *bd*. *Biochemical and Biophysical Research Communications, 355*(1), 97-102.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., *et al*. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 7*(5), 335-336.

Chirife, J., & Buera, P. (1994). Water activity, glass transition and microbial stability in concentrated/ semimoist food systems. *Journal of Food Science, 59*(5), 921-927.

Chou, C.-C., & Ling, M.-Y. (1998). Biochemical changes in soy sauce prepared with extruded and traditional raw materials. *Food Research International, 31*(6), 487-492.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics, 21*(18), 3674-3676.

Daniel, R. (2005). The metagenomics of soil. *Nature Reviews Microbiology, 3*(6), 470-478.

Derrick, S., & Large, P. J. (1993). Activities of the enzymes of the Ehrlich pathway and formation of branched-chain alcohols in *Saccharomyces cerevisiae* and *Candida utilis* grown in continuous culture on valine or ammonium as sole nitrogen source. *Journal of General Microbiology, 139*(11), 2783-2792.

Dunbar, J., White, S., & Forney, L. (1997). Genetic diversity through the looking glass: Effect of enrichment bias. *Applied and Environmental Microbiology, 63*(4), 1326-1331.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461.

Edwards, R., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D., *et al.* F. (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics, 7*(1), 57.

Eisen, J. A. (2007). Environmental shotgun sequencing: Its potential and challenges for studying the hidden world of microbes. *PLoS Biology, 5*(3), e82.

Farber, J., & Peterkin, P. (1991). *Listeria monocytogenes*, a food-borne pathogen. *Microbiological Reviews, 55*(3), 476.

Fukushima, D. (2004). Industrialization of fermented soy sauce production centering around Japanese shoyu. In Steinkraus, K. H. (Ed.), *Food Science and Technology* (pp. 1-88). New York, USA: Marcel Dekker.

Gao, X., Cui, C., Ren, J., Zhao, H., Zhao, Q., & Zhao, M. (2011). Changes in the chemical composition of traditional Chinese-type soy sauce at different stages of manufacture and its relation to taste. *International Journal of Food Science & Technology, 46*(2), 243-249.

Ghose, T., & Bandyopadhyay, K. (1980). Rapid ethanol fermentation in immobilized yeast cell reactor. *Biotechnology and Bioengineering, 22*(7), 1489-1496.

Gilbert, J. A., & Dupont, C. L. (2011). Microbial Metagenomics: Beyond the Genome. *Annual Review of Marine Science, 3*(1), 347-371.

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S. *et al*. (2006). Metagenomic analysis of the human distal gut microbiome. *Science, 312*(5778), 1355-1359.

Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R. *et al*. (2002). Isolation of antibiotics Turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology, 68*(9), 4301-4306.

Giraffa, G. (2004). Studying the dynamics of microbial populations during food fermentation. *FEMS Microbiology Reviews, 28*(2), 251-260.

Goenrich, M., Bartoschek, S., Hagemeier, C. H., Griesinger, C., & Vorholt, J. A. (2002). A glutathione-dependent formaldehyde-activating enzyme (Gfa) from *Paracoccus denitrificans* detected and purified via two-dimensional proton exchange NMR spectroscopy. *Journal of Biological Chemistry, 277*(5), 3069-3072.

Graham, L. A., Flannery, A. R., & Stevens, T. H. (2003). Structure and assembly of the yeast V-ATPase. *Journal of Bioenergetics and Biomembranes, 35*(4), 301-312.

Hamada, T., Sugishita, M., & Motai, H. (1990). Contributions of immobilized and free cells of salt-tolerant *Zygosaccharomyces rouxii* and *Candida versatilis* to the production of ethanol and 4-ethylguaiacol. *Applied Microbiology and Biotechnology, 33*(6), 624-628.

Hamberger, A., Horn, M. A., Dumont, M. G., Murrell, J. C., & Drake, H. L. (2008). Anaerobic consumers of monosaccharides in a moderately acidic fen. *Applied and Environmental Microbiology, 74*(10), 3112-3120.

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews, 68*(4), 669-685.

Handelsman, J., Tiedje, J., Alvarez-Cohen, L., Ashburner, M., Cann, I., Delong, E. *et al*. (2007). *The new science of metagenomics: Revealing the secrets of our microbial planet*. Washington, DC: National Academy of Sciences.

Haruta, S., Ueno, S., Egawa, I., Hashiguchi, K., Fujii, A., Nagano, M. *et al*. (2006). Succession of bacterial and fungal communities during a traditional pot fermentation of rice vinegar

assessed by PCR-mediated denaturing gradient gel electrophoresis. *International Journal of Food Microbiology, 109*(1), 79-87.

Hayashi, A., & Kimoto, K. (2007). Nicotianamine preferentially inhibits angiotensin I-converting enzyme. *Journal of Nutritional Science and Vitaminology, 53*(4), 331-336.

Hong, S., Bunge, J., Leslin, C., Jeon, S., & Epstein, S. S. (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME, 3*(12), 1365-1373.

Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N., & Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Research, 21*(9), 1552-1560.

Huson, D. H., Richter, D. C., Mitra, S., Auch, A. F., & Schuster, S. C. (2009). Methods for comparative metagenomics. *BMC Bioinformatics, 10*(Suppl 1), S12.

Hutkins, R. W. (2006). *Microbiology and technology of fermented foods*. Ames, Iowa, USA: Blackwell Publishing.

Hyatt, D., LoCascio, P. F., Hauser, L. J., & Uberbacher, E. C. (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics, 28*(17), 2223-2230.

Jan, G., Rouault, A., & Maubois, J.-L. (2000). Acid stress susceptibility and acid adaptation of *Propionibacterium freudenreichii* subsp. *shermanii*. *Le Lait, 80*(3), 325-336.

Jernberg, C., Löfmark, S., Edlund, C., & Jansson, J. K. (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *ISME, 1*(1), 56-66.

Jung, J. Y., Lee, S. H., Kim, J. M., Park, M. S., Bae, J.-W., Hahn, Y. *et al*. (2011). Metagenomic analysis of kimchi, a traditional Korean fermented food. *Applied and Environmental Microbiology, 77*(7), 2264-2274.

Kalyuzhnaya, M. G., Lapidus, A., Ivanova, N., Copeland, A. C., McHardy, A. C., Szeto, E. *et al*. (2008). High-resolution metagenomics targets specific functional types in complex microbial communities. *Nature Biotechnology, 26*(9), 1029-1034.

Kandler, O. (1983). Carbohydrate metabolism in lactic acid bacteria. *Antonie Van Leeuwenhoek, 49*(3), 209-224.

Kataoka, S. (2005). Functional effects of Japanese style fermented soy sauce (shoyu) and its components. *Journal of Bioscience and Bioengineering, 100*(3), 227-234.

Kembel, S. W., Wu, M., Eisen, J. A., & Green, J. L. (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biology, 8*(10), e1002743.

Kennedy, J., Flemer, B., Jackson, S. A., Lejon, D. P., Morrissey, J. P., O'Gara, F., & Dobson, A. D. (2010). Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. *Marine Drugs, 8*(3), 608-628.

Kim, T. W., Lee, J.-H., Park, M.-H., & Kim, H.-Y. (2010). Analysis of bacterial and fungal communities in Japanese- and Chinese- fermented soybean pastes using nested PCR–DGGE. *Current Microbiology, 60*(5), 315-320.

Kim, Y. S., Kim, M. C., Kwon, S. W., Kim, S. J., Park, I. C., Ka, J. O., & Weon, H. Y. (2011). Analyses of bacterial communities in meju, a Korean traditional fermented soybean bricks, by cultivation-based and pyrosequencing methods. *Journal of Microbiology, 49*(3), 340-348.

Kobayashi, M. (2005). Immunological functions of soy sauce: Hypoallergenicity and antiallergic activity of soy sauce. *Journal of Bioscience and Bioengineering, 100*(2), 144-151.

Kohn, A. (1960). Lysis of frozen and thawed cells of *Escherichia coli* by lysozyme and their conversion into spheroplasts. *Journal of Bacteriology, 79*(5), 697-706.

König, H., & Fröhlich, J. (2009). Lactic acid bacteria. In König, H., Fröhlich, J & Unden, G. (Eds.) *Biology of microorganisms on grapes, in must and in wine* (pp. 3-29) Verlag, Berlin Heidelberg, Germany: Springer.

Kurtzman, C. (2011) *Saturnispora.* In Kurtzman, C., Fell, J. W., & Boekhout, T. (Eds.). *The Yeasts: A Taxonomic Study* (5th ed.) (Vol. 2) (pp. 765-772). USA: Elsevier.

LeCleir, G. R., Buchan, A., & Hollibaugh, J. T. (2004). Chitinase gene sequences retrieved from diverse aquatic habitats reveal environment-specific distributions. *Applied and Environmental Microbiology, 70*(12), 6977-6983.

Lee, K. S., Lee, Y. B., Lee, D. S., & Chung, S. K. (2006). Quality evaluation of Korean soy sauce fermented in Korean earthenware (Onggi) with different glazes. *International Journal of Food Science and Technology, 41*(10), 1158-1163.

Leroy, F., & De Vuyst, L. (2004). Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends in Food Science & Technology, 15*(2), 67-78.

Li, X.-R., Ma, E.-B., Yan, L.-Z., Meng, H., Du, X.-W., Zhang, S.-W., & Quan, Z.-X. (2011). Bacterial and fungal diversity in the traditional Chinese liquor fermentation process. *International Journal of Food Microbiology, 146*(1), 31-37.

Liu, M., Nauta, A., Francke, C., & Siezen, R. J. (2008). Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. *Applied and Environmental Microbiology, 74*(15), 4590-4600.

Luh, B. (1995). Industrial production of soy sauce. *Journal of Industrial Microbiology & Biotechnology, 14*(6), 467-471.

Manichanh, C., Chapple, C. E., Frangeul, L., Gloux, K., Guigo, R., & Dore, J. (2008). A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. *Nucleic Acids Research, 36*(16), 5180-5188.

Marilley, L., & Casey, M. (2004). Flavours of cheese products: Metabolic pathways, analytical tools and identification of producing strains. *International Journal of Food Microbiology, 90*(2), 139-159.

Masuda, S., Yamaguchi, H., Kurokawa, T., Shirakami, T., Tsuji, R. F., & Nishimura, I. (2008). Immunomodulatory effect of halophilic lactic acid bacterium *Tetragenococcus halophilus* Th221 from soy sauce moromi grown in high-salt medium. *International Journal of Food Microbiology, 121*(3), 245-252.

Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C. *et al*. Food-related illness and death in the United States. *Emerging Infectious Diseases, 5*(5), 607.

Moat, A. G., Foster, J. W., & Spector, M. P. (2003). In *Microbial Physiology* (pp. 412-433) New York, USA: Wiley-Liss.

Nikolausz, M., Sipos, R., Revesz, S., Szekely, A., & Marialigeti, K. (2005). Observation of bias associated with re-amplification of DNA isolated from denaturing gradient gels. *FEMS Microbiology Letters, 244*(2), 385-390.

Nishi, T., & Forgac, M. (2002). The vacuolar (H+)-ATPases - nature's most versatile proton pumps. *Nature Reviews Molecular Cell Biology, 3*(2), 94-103.

Nissen, T. L., Kielland-Brandt, M. C., Nielsen, J., & Villadsen, J. (2000). Optimization of ethanol production in *Saccharomyces cerevisiae* by metabolic engineering of the ammonium assimilation. *Metabolic Engineering, 2*(1), 69-77.

Noda, F., Hayashi, K., & Mizunuma, T. (1980). Antagonism between osmophilic lactic acid bacteria and yeasts in brine fermentation of soy sauce. *Applied and Environmental Microbiology, 40*(3), 452-457.

Noda, F., Hayashi, K., & Mizunuma, T. (1982). Influence of pH on inhibitory activity of acetic acid on osmophilic yeasts used in brine fermentation of soy sauce. *Applied and Environmental Microbiology, 43*(1), 245-246.

O'toole, D. K. (1997). The role of microorganisms in soy sauce production. *Advances in Applied Microbiology, 45*, 87-152.

Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P. *et al*. (2008). KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Research, 36* (suppl 2), W423-W426.

Paul Ross, R., Morgan, S., & Hill, C. (2002). Preservation and fermentation: Past, present and future. *International Journal of Food Microbiology, 79*(1), 3-16.

Pettengill, J. B., McAvoy, E., White, J. R., Allard, M., Brown, E., & Ottesen, A. (2012). Using metagenomic analyses to estimate the consequences of enrichment bias for pathogen detection. *BMC Research Notes, 5*, 378.

Polz, M. F., & Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology, 64*(10), 3724-3730.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: A comprehensive online resource for quality checked and aligned

ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research, 35*(21), 7188-7196.

Quigley, L., O'Sullivan, O., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., & Cotter, P. D. (2012). High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. *Applied and Environmental Microbiology, 78*(16), 5717-5723.

Rattanachaikunsopon, P., & Phumkhachorn, P. (2010). Lactic acid bacteria: Their antimicrobial compounds and their uses in food production. *Annals of Biological Research, 1*, 218-228.

Rosin, F. M., Watanabe, N., & Lam, E. (2005). Moonlighting vacuolar protease: Multiple jobs for a busy protein. *Trends in Plant Science, 10*(11), 516-518.

Scholz, M. B., Lo, C. C., & Chain, P. S. (2012). Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. *Current Opinion in Biotechnology, 23*(1), 9-15.

Shah, N., Tang, H., Doak, T. G., & Ye, Y. (2011). Proceeding from Pacific Symposium on Biocomputing: *Comparing Bacterial Communities Inferred from 16S rRNA Gene Sequencing and Shotgun Metagenomics.* Fairmont Orchid, HI: PSB.

Simon, C., & Daniel, R. (2009). Achievements and new knowledge unraveled by metagenomic approaches. *Applied Microbiology and Biotechnology, 85*(2), 265-276.

Simon, C., & Daniel, R. (2011). Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology, 77*(4), 1153-1161.

Smit, G., Smit, B. A., & Engels, W. J. (2005). Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiology Reviews, 29*(3), 591-610.

Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research, 33*, W465-467.

Suenaga, H., Ohnuki, T., & Miyazaki, K. (2007). Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environmental Microbiology, 9*(9), 2289-2297.

Tanaka, Y., Watanabe, J., & Mogi, Y. (2012). Monitoring of the microbial communities involved in the soy sauce manufacturing process by PCR-denaturing gradient gel electrophoresis. *Food Microbiology, 31*(1), 100-106.

Tanasupawat, S., Thongsanit, J., Okada, S., & Komagata, K. (2002). Lactic acid bacteria isolated from soy sauce mash in Thailand. *Journal of General and Applied Microbiology, 48*(4), 201-209.

ter Schure, E. G., Flikweert, M. T., van Dijken, J. P., Pronk, J. T., & Verrips, C. T. (1998). Pyruvate decarboxylase catalyzes decarboxylation of branched-chain 2-oxo acids but is not essential for fusel alcohol production by *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology, 64*(4), 1303-1307.

Torsvik, V., Goksøyr, J., & Daae, F. L. (1990). High diversity in DNA of soil bacteria. *Applied and Environmental Microbiology, 56*(3), 782-787.

Tsuchiya, H., Sato, M., & Watanabe, I. (1999). Antiplatelet activity of soy sauce as functional seasoning. *Journal of Agricultural and Food Chemistry, 47*(10), 4167-4174.

van der Sluis, C., Stoffelen, C. J., Castelein, S. J., Engbers, G. H., ter Schure, E. G., Tramper, J., & Wijffels, R. H. (2001). Immobilized salt-tolerant yeasts: application of a new polyethylene-oxide support in a continuous stirred-tank reactor for flavour production. *Journal of Biotechnology, 88*(2), 129-139.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A. *et al*. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science, 304*(5667), 66-74.

Von Wright, A. & Axelsson, L. (2004). Lactic acid bacteria: An introduction. In Salminen, S., Von Wright, A. & Ouwehand, A. (Eds.), *Lactic Acid Bacteria: Microbiology and Functional Aspects* (pp. 2-16). New York, USA: Marcel Dekker.

Yamada, T., Letunic, I., Okuda, S., Kanehisa, M., & Bork, P. (2011). iPath 2.0: Interactive pathway explorer. *Nucleic Acids Research, 39* (suppl 2), W412-W415.

Ye, Y., Choi, J. H., & Tang, H. (2011). RAPSearch: A fast protein similarity search tool for short reads. *BMC Bioinformatics, 12*, 159.

Ye, Y., & Doak, T. G. (2009). A parsimony approach to biological pathway reconstruction/ inference for genomes and metagenomes. *PLoS Computational Biology, 5*(8), e1000465.

Yokotsuka, T., & Sasaki, M. (1997). Fermented protein foods in the Orient: Shoyu and miso in Japan. In Woods, B.J.B. (Ed.), *Microbiology of Fermented Foods* (pp. 351-415). New York, USA: Springer.

Yong, F., & Wood, B. (1974). Microbiology and biochemistry of soy sauce fermentation. *Advances in Applied Microbiology, 17*, 157-194.

Yvon, M., & Rijnen, L. (2001). Cheese flavour formation by amino acid catabolism. *International Dairy Journal, 11*(4), 185-201.

Zhao, J., Dai, X., Liu, X., Chen, H., Tang, J., Zhang, H., & Chen, W. (2009). Changes in microbial community during Chinese traditional soybean paste fermentation. *International Journal of Food Science & Technology, 44*(12), 2526-2530.